Modern Environmental Science and Engineering (ISSN 2333-2581)
September 2018, Volume 4, No. 9, pp. 872-881
Doi: 10.15341/mese(2333-2581)/09.04.2018/011
Academic Star Publishing Company, 2018
www.academicstar.us

Automatic Environmental Zoning with Self-Organizing Maps

Marcos A. Santos da Silva[1], Renato J. S. Maciel[2], Leonardo N. Matos[3], and Márcia H. G. Dompieri[4]

1. Embrapa Tabuleiros Costeiros, Brazilian Agricultural Research Corporation, Brazil

2. Embrapa Informática Agropecuária, Brazilian Agricultural Research Corporation, Brazil

3. Department of Computer Science, Federal University of Sergipe, Brazil

4. Embrapa Territorial, Brazilian Agricultural Research Corporation, Brazil

Abstract: This article presents the application of the Self-Organizing Maps (SOM) as an exploratory tool for automatic environmental zoning by combining the handle of categorical data and the other for automatic clustering. The SOM online learning algorithm had been chosen to treat categorical data by using the dot product method and the Sorense-Dice binary similarity coefficient. To automatically perform a spatial clustering, an adaptation of the automatic clustering Costa-Netto algorithm had been also proposed. The correspondence analysis had been used to examine the profiles of each homogeneous zones. To explore the approach it has been performed the environmental zoning of the Alto Taquari River Basin, Brazil, using as input data a set of thematic maps. The results indicate the applicability of the approach to perform the exploratory environmental zoning.

Key words: artificial neural network, exploratory spatial analysis, similarity coefficients, correspondence analysis, Alto Taquari river basin

1. Introduction

The power demand in agriculture is characterized by varied production processes and their influence on consumption. The structural development, the mechanization and automation, the condition and the age of the husbandry- and process-technology are factors which have an effect on the power demand of the agricultural operation. The cost factor “energy” is not to be underestimated in the management. Continuously rising energy prices and the increasing power demand in the production show a considerable monetary charge for agricultural farms.

Territorial zoning has been showing high strategic and tactical potential to public managers, research institutions, rural entrepreneurs, universities and other institutions concerned with spatial planning. Spatial or regional planning has as general goal a better use of the geographic space from multisource valuable information and knowledge. Despite improvements in GIS and remote sensing, spatial planning is still facing many challenges, mainly concerned about spatial data processing and interpretation [1, 2]. In fact, many spatial planning problems, such as ecological-economic zoning (also known as environmental zoning), are complex and context dependent, thus very difficult to generalize. In this context, computational geospatial data exploration tools based on machine learning techniques play a central role as a support for spatial planning.

Several types of territorial zoning exist, such as urban, industrial, agro-ecological, ecological, economic and environmental. Each type of territorial zoning adopts its own methods to detect homogeneous zones using quantitative and qualitative criteria. In

Corresponding author: Marcos Aurélio Santos da Silva, Ph.D. Candidate in Artificial Intelligence; research areas/interests: geocomputation and computational social science. E-mail: marcos.santos-silva@embrapa.br.

general, due to the complexity of the zoning process, it is necessary to gather a bunch of experts to interpret the geospatial data, including remote sensing data, and to design the homogeneous zones.

Nevertheless, due to the increasing demand for different zoning (e.g., resource management, policy making, contingency plans) it is worth make use of machine learning techniques to speed up this process of territorial zoning, mainly environmental zoning in developing countries where it is observed a growing in agricultural activities with a huge pressure over the environment [3].

One challenging task in this endeavor, territorial zoning, is the interpretation, at the same time, of quantitative and qualitative (e.g., thematic maps) data. There are two ways to approach this type of hybrid data (categorical and numerical), all data are transformed into categorical or numerical, or the algorithms can be adapted to treat at the same time both types of data. In general, these data are complex, so they have a high degree of uncertainty, and redundancy; they present a structural (in terms of topology) complex data arrangement and a high rate of incorrect data (noise). In this work, all quantitative data were transformed into categorical to facilitate the data processing and interpretation.

Thus, it is essential to use and develop methods to process categorical data considering issues such to capture the complexities inherent in the data gathered and used in territorial zoning. One of these methods is the Kohonen Self-Organizing Map (SOM) Artificial Neural Network [4, 5], which has the potential for cluster analysis and visualisation of small and large datasets [6]. As stated by Goodchild (2008) [7] "... SOM provides more than a single addition to the spatial analytic toolbox, but instead reflects an entirely new paradigm for ESDA [Exploratory Spatial Data Analysis] and spatial data mining". Despite the wide use of the SOM for spatial analysis, there are few works on the analysis of these types of geospatial data.

The objective of this study has been to investigate an adaptation of the SOM neural network for automatic environmental zoning by means of a heuristic and automatic process for clustering categorical spatial data (thematic maps). The proposed method had been demonstrated for the environmental zoning of the Alto Taquari River Basin, Brazil [8], taking as input dataset thematic maps of the region. The results were compared with those obtained by Silva and Santos (2011) [9], who used a combination of statistical methods to design an environmental zoning of this region.

Section 2 shortly introduces the Self-Organizing Map. Section 3 provides a brief review on the use of Self-organizing Maps for exploratory spatial analysis and categorical data classification. Section 4 presents the proposed approach. Section 5 describes the case study, ecological-economic zoning of the Alto Taquari River Basin, Brazil. Section 6 shows the results and discussions. Finally, Section 7 presents the general conclusions.

2. Literature Review

2.1 SOM for Spatial Data Exploration

Since the early 90's, many works have been focusing on the application of self-organizing maps to geospatial data exploration, including clustering and visualization by data ordering [1, 6, 10-20]. In general, these works explore the use of a two-dimensional standard SOM for unsupervised geo-spatial data exploration. Some works focus on the investigation of spatial constraints of the geospatial data [21] or performs regionalization [17], [22]. Despite the number of papers which use the self-organizing map as a geospatial data clustering or visualization method, there are few works about automatic zoning using categorical data (e.g., thematic maps) [1].

Sadek, Lima & Adami (2017) [1] combined a SOM artificial neural network as a vector quantization tool with k-means clustering technique (associated with a clustering index, Davies-Bouldin) to perform an automatic ecological-economic zoning. But, instead of treat categorical data directly, they transformed categorical data into numerical and applied the standard SOM's learning algorithm. In this implementation, each neuron was considered a center of class input to the k-means.

For clustering of numerical geospatial data aggregated by area [23] used an automatic clustering method based on the graph partition [24], by means of a standard SOM and two clustering validity indexes, Davies-Bouldin [25] and CDbw [23, 26-30]. This approach showed to be robust in real applications [23], [26-28] and it had been chosen as part of the proposed method.

2.2 SOM for Categorical Data

The most common use of SOM is for topological ordering. This means that the proximity in the neural map will obey the proximity in the feature space of the original data. This property can be used for clustering and visualization by projecting each data identification on the neuron that corresponds to the Best Match Unit (BMU). This topological ordering is analogous to the principal component analysis, for continuous data, and to the correspondence analysis, for categorical data [4].

Following this approach, Cottrell and Letrémy (2005) [31] proposed three variants of SOM to treat categorical data, the Kohonen Multiple Correspondence Analysis (KMCA) to explore relations among modalities, KMCA with individuals that uses the same trained SOM to project modalities and individuals, not simultaneously, and the Kohonen Disjunctive table that treats individuals and modalities at the same time.

In general, the use of SOM to perform multivariate exploratory categorical data analysis depends on the transformation of data. Some authors transformed the categorical data into a complete disjunctive table and then adapted the SOM algorithm to treat this table [31-33].

López-Rubio (2010) [34] used a stochastic approximation learning rule into a SOM to treat ordinary categorical data set. [35] developed two SOM variants to handle a data set composed of a mixed (numerical and categorical) hierarchical data. Jiao, Liu, and Zou (2011) [36] designed a new distance measure and a tree-growing adaptation method into the SOM for transactional data. These authors also used the U-matrix for data visualization. Yang, Hung, and Chen (2012) [37] used symbolic neurons and a suppression learning method to process heterogeneous data, including nominal, interval and ordinal types of data. Jiao, Liu, and Zou (2011) [36] used spatial (points) and non-spatial attributes, at the same time, to perform a dual clustering using a SOM variant to explore the spatial distribution of non-spatial data for urban land price regionalization. Coso et al. (2015) [38] directly process categorical data without any data transformation, but adapted the SOM batch learning process, including a frequency table to treat categorical data.

Lourenço, Lobo & Bação (2004) [33] used two methods to treat categorical data: the hard logic and the dot product. In the hard logic method, although the network weights w_t have real values, they are interpreted as binary logical values using a real number t as a threshold. For example, if $w_t > t$, then its value is assumed to be one, otherwise it is zero. In the dot product method, a weight vector is interpreted as a vector of probabilities, where each component value represents the probability of it being equal to one. In both situations, during the learning process, the weights have real values and are updated in the same way as in the original algorithm.

In this work it had been adopted the standard online learning SOM, one and bidimensional, because there exists a huge number of successful applications with geospatial as showed above. The SOM had been used to a clustering problem, that is an automatic environmental zoning, using the data partition solution proposed by Costa and Netto (2003) [24] that was applied with successfully using numerical geospatial data by Silva (2010) [26], Silva (2015) [28] where it is not necessary to establish a priori number of clusters (homogeneous zones). The environmental zoning will be performed using only categorical data (thematic maps) because this simplifies the entire process and facilitates the interpretation of results. The categorical data will be treated as binary data, instead of transforming it into numerical values, so the solution could be generalized for any other type of territorial zoning. The SOM will be adapted to treat these binary data using the simplest solution, the hard logic solution proposed by Lourenço, Lobo, and Bação (2004) [33].

3. Material and Methods

3.1 Self-Organizing Maps

The Self-Organizing Map is an unsupervised artificial neural network composed of a set of connected code vectors, also called network weights,

$$\mathbf{w}_p = [w_{1p}, w_{2p}, \ldots, w_{Mp}]^{\mathrm{T}}, \qquad p = 1..S,$$

where S is the total number of code vectors, organized in a one or two dimensional regular grid that approximates the input data, $\mathbf{X}_{N\times M}$, by a vector quantization strategy based on a learning process that comprises three phases: competitive, cooperative and adaptive. At the end of the learning process, each code vector is associated with a subset of the input data. The equation below shows the adaptive phase of the online (stochastic) learning process at time t.

$$\mathbf{w}_p(t+1) = \mathbf{w}(t) + \alpha(t)\gamma(t)\tau(\mathbf{w}_p - \mathbf{x}_i)$$

where $\alpha(t)$ and $\gamma(t)$ represents monotonically decreasing functions of learning rate and neighborhood between code vectors $\mathbf{w}_i$ and $\mathbf{w}_j$, and $\tau\ (\mathbf{w}_p - \mathbf{x}_\imath)$ represents a distance measure between $\mathbf{w}_p$ and $\mathbf{x}_\imath$. The vector $\mathbf{x}_\imath$ represents the ıth input data. More detail about the standard SOM can be found in Refs. [4, 5].

The experiments conducted in this project used two types of SOM, one-dimensional and two-dimensional, with hexagonal lattices, varying the neighborhood initial radius. The number of iterations for the learning phase had been fixed to one thousand.

Although the SOM has not been designed for automatic clustering, many studies have combined it with other methods to do so. For automatic clustering using SOM we have the work of Costa and Netto (2003) [24] that used mathematical morphology to automatically segment the U-matrix [39]. Another way to perform data clustering is to use the SOM for data ordering and after that uses the code vectors as input for a statistical clustering technique like k-means or the agglomerative hierarchical algorithm [4].

To proceed the experiments, it has been used the SOMCode software and the GIS Spring.

It is important to notice that despite the number of heuristic parameters needed to set up a SOM, it has been shown that this artificial neural network maintains the similar topological order for the same data even using different initial values of SOM's constants and initial parameters [4, 5].

3.2 The Proposed Method

In this paper the SOM online learning algorithm was adapted to treat categorical data [33, 40] and it had been used in pair with the SOM's graph partition segmentation, the Costa-Netto algorithm [24, 26] to, automatically, cluster thematic maps from the Alto Taquari River Basin, Mato Grosso do Sul, Brazil. Some binary similarity measure had been tested in the experiments.

The proposed approach for automatic environmental zoning through the classification of categorical data using SOM can be summarized as follows (Fig. 1): 1) A routine for extracting features from thematic maps is applied; 2) The adapted Self-Organizing map is trained using this binary data as input; 3) The adapted Costa-Netto algorithm is applied to automatic cluster the neural network; 4) The classified thematic map is generated using the labeled SOM; 5) Interpretation of the profiles of each environmental zone using Correspondence Analysis [41, 42].

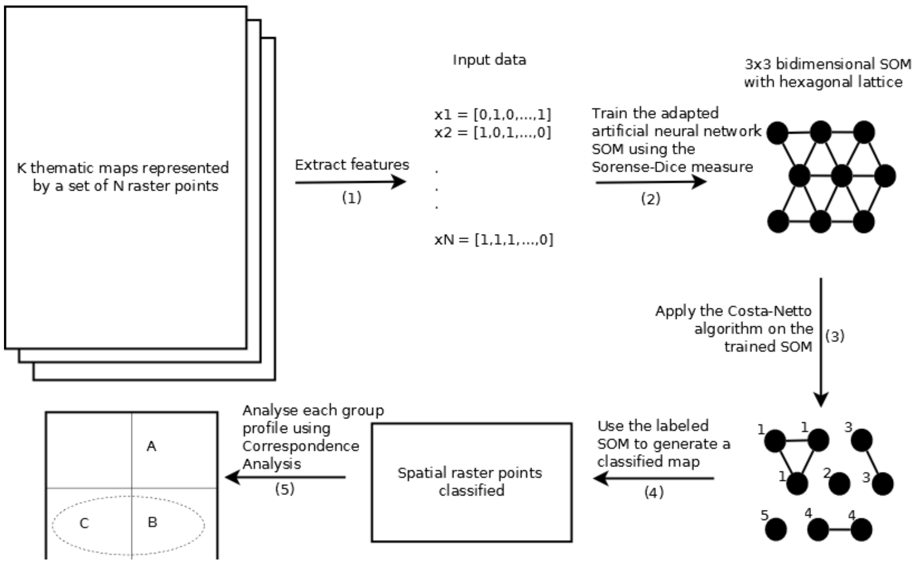


Fig. 1 Scheme of the proposed method for investigation of environmental zoning using thematic maps.

Algorithm 1 Costa Netto Algorithm

Require: m — the number of neurons

Require: n — the number of input patterns

Require: $G = \{V, A\}$ — graph based on the neural network, where V is the neuron set and A is the set of arcs

Require: $H(\imath)$ — activity function of neuron i, $i \in \{1, \ldots, m\}$

Require: ω — $0.1 \leq \omega < 0.6$

1: $H_{min} \leftarrow \omega(n/m)$ — minimum allowed value for $H(\imath)$

2: **for** pair of adjacent neurons i and j **do**

3: Compute $d(\mathbf{w}_\imath ; \mathbf{w}_j)$ — the distance between the weights of neurons I and j

4: Compute $\bar{d}_{i,j}$ — mean distance between the i's and j's neighbors

5: Compute c_i — centroid of each neuron

6: **for all** pair of adjacent neurons i and j **do**

7: **if** $(d(\mathbf{w}_i ; \mathbf{w}_j) > 2\bar{d}_{i,j})$ **or**

8: (($H(\imath) < H_{min}/2$ **or** $H(j) < H_{min}/2$) **and** ($H(i) = 0$ **or** $H(j) = 0$)) **or**

9: ($\mathbf{c}_i > 2d(\mathbf{w}_i; \mathbf{w}_j)$ or $\mathbf{c}_j > 2d(\mathbf{w}_i ; \mathbf{w}_j)$) **then**

10: Remove arc($\imath$, j) from A

11: A distinct label is assigned to each set of connected neurons to identify the group.

3.2.1 Feature Extraction from Thematic Maps

Let's consider a set of K thematic maps represented by a set of N raster points, each map has m_k classes, if $M = \sum_{k=1}^{K} m_k$ is the total number of classes, each input vector will be represented by the vector $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots x_{iM}]^T$, $i = 1..N$, where $x_{ij} \in \{0,1\}$, and x_{ij} assumes 1 if the class is present, 0 otherwise.

3.2.2 Adapted SOM and Costa-Netto Algorithm

The original SOM algorithm had been adapted to treat binary data using the hard logic proposed by Lourenço, Lobo, and Bação (2004) [33].

The Costa-Netto algorithm is defined as follows, according to Costa and Netto (2003) [24], Silva et al. (2010) [26], and Silva et al. (2015) [28]. The distance measure was changed following the solution proposed to the learning process of the SOM, the hard logic measure. The algorithm had been adapted for categorical data and it used some binary distance to measures to evaluate the proximity among code vectors. The code has been implemented in the SOMCode Project.

At the end of the process, only connected nodes representing different groups shall remain. The algorithm had been adapted for categorical data and it used some binary distance to measures to evaluate the proximity among code vectors. The code has been implemented in the SOMCode Project.

3.2.3 Similarity Measures for Categorical Data

Another important point regarding the use of SOM for categorical data is the type of binary distance used. Table 1 presents the binary distances analyzed to perform the training and segmentation algorithms during the experiments.

When comparing two binary vectors **x** and **y**, a = number of times that $x_i = 1$ and $y_i = 1$; b = number of times that $x_i = 0$ and $y_i = 1$; c = number of times that $x_i = 1$ and $y_i = 0$; and d = number of times that $x_i = 0$ and $y_i = 0$. But, when considering the feature vectors generated from K thematic maps, the number of positive co-occurrence a varies from 0 to K at most, and $b = c = K - a$, thus, in this situation, for instance, Sorense-Dice will be equal to a/K. If M increases when compared to the number of thematic maps, K, d increases and dominates the binary coefficients that includes this variable in its calculus. Therefore, it had been decided to exclude from this study these coefficients.

Lourenço, Lobo & Bação (2004) [33] performed a specific study on the use of these similarity coefficients for categorical data and their application in SOM and they concluded that they are very sensitive when used for clustering.

3.2.4 Analysis of Zones Profiles

The final environmental zoning map can be analyzed through the frequency distribution table (contingency table) of each class within the homogeneous zones (zones profiles). Different territorial zonings can be compared even if they have non-identical numbers of clusters. According to Henriques, Bação, and Lobo (2012) [41], the spatial dependence among the geospatial data limits the use of inferential statistics.

Table 1 Similarity coefficients explored as a function of *a*, *b*, *c* and *d*. Source: Bação, Lobo & Painho (2005).

Coefficients	Equation	Equation for thematic maps
Sorence-Dice	2a/(2a+b+c)	a/K
Anderberg	a/(a+2(b+c))	$a/(4K-3a)$
Ochiai	a/(a+b)(a+c)	a/K^2

The contingency table can also be studied by the Correspondence Analysis. This descriptive statistical method decomposes the total variance, in relation to the center of the cloud data, into orthogonal factors, so the data can be analyzed in terms of their relative position on these factors. This technique generates substantial information about the zones' profiles, which pemits: i) the display of the cloud data into a two-dimensional graph map; ii) the calculation of the chi-square distance of each zone profile from the center of the data; iii) the evaluation of the mass of each zone and class; iv) the calculation of the percentage of the inertia explained by each zone or class; v) a measurement estimation of the quality of the zone or class projection of the factor dimensions, based on the contribution of one factor to the square of the distance of the zone or class from the center of the data.

3.3 The Case Study: Environmental Zoning of the Alto Taquari River Basin

The Alto Taquari River Basin (BAT) is adjacent to the Pantanal, in Mato Grosso do Sul, Brazil. This area covers 28,046 km^2. It presents a demographic density of 2.53 inhabitants/km^2, and has extensive livestock farming and agriculture (grains) as the main economic activities [8]. It has been used five thematic maps in the experiments (Table 2), covering five dimensions: environmental, infrastructure, economic aspects, population dynamics and living conditions of the population [9].

These five thematic maps were generated from forty indicators chosen to identify the pressure, the state and the response on the BAT watershed based on the OECD (1993) methodology. These maps are in the 1:250000 scale, the raster files has 1085 rows and 817 columns and the total number of feature vectors is 448724. The feature vectors were extracted from these five maps using the method explained in subsection 4.1. Therefore, each vector has twenty variables related to the categories in the maps (see Table 2). Although, each thematic map represents a set of ordinary classes, the data set had been treated as categorical.

Table 2 Thematic Maps of the Alto Taquari River Basin, Ms, Brazil.

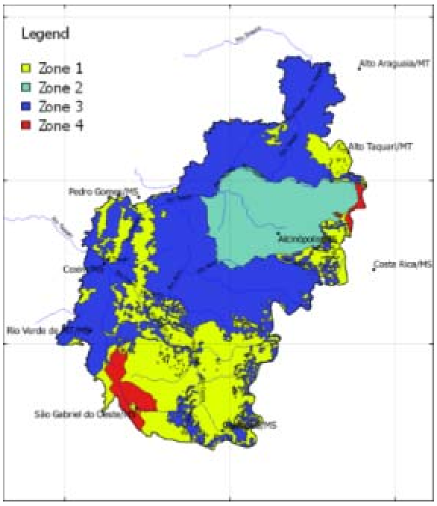
Thematic map	Classes	Labels
Economic aspects	Low, medium and high quality of the economic aspect index	EA1, EA2 and EA3
Environmental dimension	Eight groups organized in the crescent order of homogeneity, from ED1 to ED8	ED1, ED2, ED3, ED4, ED5, ED6, ED7 and ED8
Infrastructure	Low, medium and high quality of the infrastructure index	IS1, IS2 and IS3
Living conditions	Low, medium and high living conditions index	LC1, LC2 and LC3
Population dynamics	Low, medium and high equilibria of the population dynamics	PD1, PD2 and PD3

4. Results and Discussion

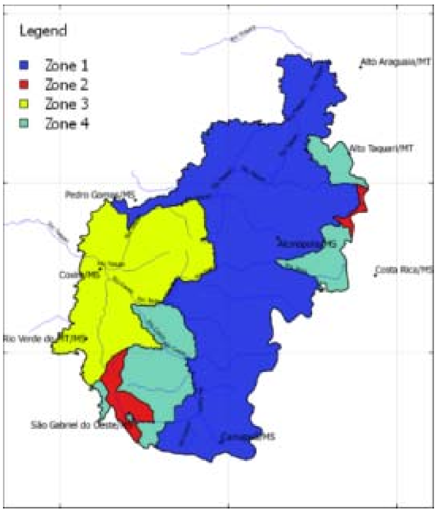
Several Self-Organizing Maps configurations were tested, including one and two dimensional neural maps, but only two of them, using the Sorense-Dice coefficient of binary similarity, generated more than two zones (clusters), a 1D SOM 1x15 and a 2D SOM 17×17. Therefore, this means that the adapted SOM for binary data is very sensitive to small changes in initial parameters.

Fig. 2(a) presents the result obtained in Ref. [9], Figs. 2(b) and 2(c) show the results obtained using the method proposed in this study. For the 1D SOM map of 1 × 15 with initial radius 3 has been obtained the same number of groups (zones) from Ref. [9]. A different result had been observed in the 17×17 2D SOM map with initial radius 8, in this case, there has been a greater influence on the environmental dimension, which explains the lower similarity in relation to the result obtained in [9].

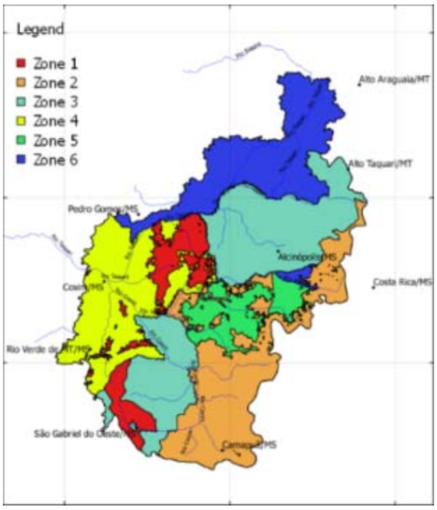
It had been applied the Correspondence Analysis over these three territorial partitions. Table 3 shows the mass, the chi-square distance to the data center and the inertia explained by each generated zone. Then, from the zones created by Silva and Santos (2011) [9] only two explains almost all inertia (ZTVila2 and ZTVila3). The one-dimensional SOM distributed the inertia almost equally among the four zones, and the



(a) Environmental zoning obtained in Ref. [9]



(b) Environmental zoning obtained in this work for a 1×15 neural network map with neighborhood radius 3



(c) Environmental zoning obtained in this work for a 17 × 17 neural network map.

Fig. 2 Comparison of results obtained in this work with the result achieved in Ref. [9].

Table 3 The mass, the chi-distance to the data center and the inertia calculated by the correspondence analysis method for each environmental zoning set of groups.

Zone/group	Mass	ChiDist	Inertia
ZTVila1	0.743479	0.199826	0.029688
ZTVila2	0.113763	1.010926	0.116262
ZTVila3	0.115641	1.291045	0.192750
ZTVila4	0.027118	1.919004	0.099863
ZTSOMUni1	0.581934	0.603867	0.212205
ZTSOMUni2	0.037206	2.469866	0.226968
ZTSOMUni3	0.226756	1.043839	0.247074
ZTSOMUni4	0.154103	1.231564	0.233736
ZTSOMBi1	0.084398	1.206881	0.122931
ZTSOMBi2	0.216021	0.880525	0.167487
ZTSOMBi3	0.275548	0.889580	0.218056
ZTSOMBi4	0.173119	1.157951	0.232127
ZTSOMBi5	0.073440	1.105970	0.089829
ZTSOMBi6	0.177473	1.078160	0.206300

two-dimensional SOM well-distributed the inertia among five zones, only the ZTSOMbi5 represents a small fraction of the total inertia. It is worth to note that the cumulative inertia of the two main dimensions of the two-dimensional SOM explains only 71.0% of the total variation, while the Vila strategy and the one-dimensional SOM explained, each one, more than 84.0%. This suggests that the two-dimensional SOM well separated the data.

The two-dimensional SOM created zones with well distributed masses and more equidistant to the center of the data cloud. The environmental zoning elaborated by Silva and Santos (2011) [9] generated one group (ZTVila1) which represents more than 74% of the total mass but are closer to the center and corresponds to a small fraction of the total inertia. The same occurs to the zone ZTSOMUni1 which concentrates a great portion of the mass, and are close to the barycenter of the data.

The graphical Correspondence Analysis (Fig. 3) had been performed on the best data partition, the two-dimensional SOM environmental zoning. It has been considered the first two principal dimensions, which explains 70.95% of the total inertia. Analyzing

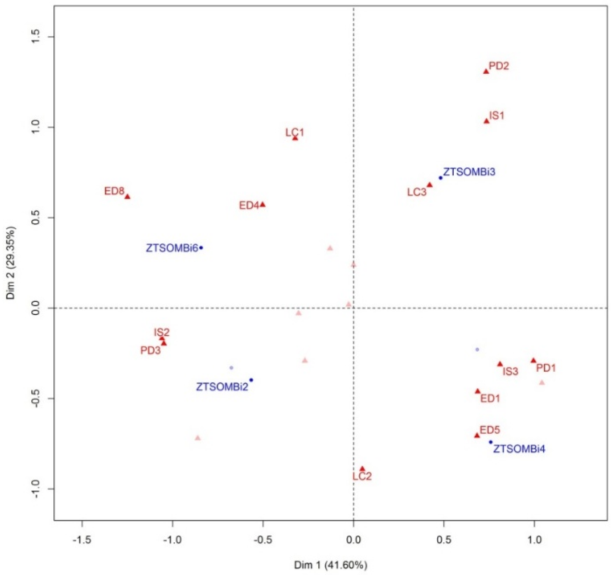


Fig 3 Correspondence analysis graph for the environmental zoning created by the two-dimensional SOM. It has been considered the zones and classes which most contributed to the two main axes.

the first axis, which explains 41.60% of the total variation and considering the four zones and the fifteen classes which most contributed to this axis, it is observed an opposition among the zones ZTSOMbi2/ZTSOMbi6 and ZTSOMbi3/ZTSOMbi4.

5. Conclusion

The two-dimensional Self-Organizing map, associated with an automatic segmentation strategy (Costa-Netto algorithm), well separated the spatial binary patterns as showed the correspondence analysis of the group’s profiles. The Sorense-Dice measure of binary similarity presented the best results. Future works may improve this calculation to take into account the differences among thematic maps with a different number of classes because if one thematic map has many classes the probability of two locations to present the same class is less than another thematic map with fewer classes. Another improvement should be the adaptation of the SOM to treat ordinary or heterogeneous data.

References

[1] L. W. R. Sadek, A. M. M. de Lima and M. Adami,

Artificial neural network for ecological-economic zoning as a tool for spatial planning, *Pesqui. Agropecuária Bras.* 52 (2017) (11) 1050-1062.

[2] G. G. Brown and N. Fagerholm, Empirical PPGIS/PGIS mapping of ecosystem services : A review and evaluation, *Ecosyst. Serv.* 13 (2015) 119-133.

[3] A. Sabatini, M. del. C. Verdiell, I. R. M. R., and M. Vidal, A quantitative method for zoning of protected areas and its spatial ecological implications, *J. Environ. Manage.* 83 (2007) (2) 198-206.

[4] T. Kohonen, Self-Organizing Maps, 3rd ed. Berlin: Springer, 2001.

[5] T. Kohonen, Essentials of the self-organizing map, *Neural Networks* 37 (2013) 52-65.

[6] P. Agarwal and A. Skupin, *Self-Organizing Maps: Applications in Geographic Information Science*, West Sussex: John Wiley & Sons, 2008.

[7] M. Goodchild, Intelligent systems for giscience: Where next? a giscience perspective, in: *Self-Organizing Maps: Applications in Geographic Information Science*, 2008, pp. 195-198.

[8] J. A. Schiavo et al., Characterization and classification of soils in the taquari river basin-pantanal region, state of mato grosso do sul, Brazil, *Rev. Bras. Ciência do Solo* 36 (2012) (3) 697-708.

[9] J. d. S. V. D. Silva and R. F. D. Santos, Estratégia metodológica para zoneamento ambiental: A experiência aplicada na Bacia Hidrográfica do Alto Rio Taquari, Campinas: Embrapa Informática Agropecuária, 2011.

[10] K. Winter and B. C. Hewitson, Self organizing maps application to census data, in: *Neural Nets: Applications in Geography*, Berlin: Springer, 1994, pp. 71-77.

[11] S. Openshaw and C. Wymer, "-Classifying and regionalizing census data, in: *Census Users Handbook*, Cambridge: GeoInformation International, 1995, pp. 239-268.

[12] S. Openshaw and I. Turton, A parallel kohonen algorithm for the classification of large spatialdatasets, *Comput. Geosci.* 22 (1996) (9) 1019-1026,.

[13] M. Takatsuka, An application of the self-organizing map and interactive 3-d visualization to geospatial data, in: *6th International Conference on GeoComputation*, 2001, pp. 24-26.

[14] M. Gahegan, M. Takatsuka, M. Wheeler, and F. Hardisty, Introducing geovista studio: Anintegrated suite of visualization and computational methods for exploration and knowledge construction in geography, *Comput. Environ. Urban Syst.* 26 (2002) (4) 267-292.

[15] J. Gorricha and V. Lobo, Improvements on the visualization of clusters in geo-referenced data using self-organizing maps, *Comput. Geosci.* 43 (2012) 177-186.

[16] N. Wang, T. W. Biggs, and A. Skupin, Visualizing gridded time series data with self-organizing maps: Anapplication to multi-year snow dynamics inthe Northern Hemisphere, *Comput. Environ. Urban Syst.* 39 (2013) 107-120.

[17] C. C. Feng, Y. C. Wang and C. Y. Chen, Combining geo-som and hierarchical clustering to explore geospatial data, *Trans. GIS* 18 (2014) (1) 125-146.

[18] E. Steiger, B. Resch, J. P. D. Albuquerque, and A. Zipf, Mining and correlating traffic events from human sensor observations with official transport data using self-organizing maps, *Transp. Res. Part C* 73 (2016) 91-104,.

[19] X. Zeng, R. He and H. Zong, Variability of Changjiang diluted water revealed by a 45-year long-term ocean hindcast and self-organizing maps analysis, *Cont. Shelf Res.*146 (2017) 37-46.

[20] D. Fei, Q. Cheng, X. Mao, F. Liu and Q. Zhou, Land use zoning using a coupled gridding-self-organizing feature maps method: A case study in China, *J. Clean. Prod.* 161 (2017) 1162-1170.

[21] G. P. Babu, Self-organizing neural networks for spatial data, *Pattern Recognit. Lett.* 18 (1997) (2) 133-142.

[22] F. Bação, V. Lobo and M. Painho, The self-organizing map, the geo-som, and relevant variants for geosciences, *Comput. Geosci.* 31 (2005) (2) 155-163.

[23] M. A. Santos da Silva, Mapas auto-organizáveis na análise exploratória de dados geoespaciais multivariados, Instituto Nacional de Pesquisas Espaciais, 2004.

[24] J. A. Costa and A. M. L. Netto, Segmentação do SOM baseada em particionamento de grafos, in: *VI Congresso Brasileiro de Redes Neurais*, 2003, pp. 451-456.

[25] D. L. Davies and D. W. Bouldin, A cluster separation measure, *IEEE Trans. Pattern Anal. Mach. IntellAnal. Mach. Intell* PAMI-1 (1979) (2) 224-227.

[26] M. A. S. da Silva, E. R. de Siqueira, O. A. Teixeira, M. G. L. Manos, and A. M. V Monteiro, Using self-organizing maps for rural territorial typology, in: *Computational Methods for Agricultural Research: Advances and Applications*, 2010, pp. 107-126.

[27] M. A. S. da Silva, E. R. de Siqueira and O. A. Teixeira, Abordagem conexionista para análise espacial exploratória de dados socioeconômicos de territórios rurais, *Rev. Econ. e Sociol. Rural* 48 (Dec. 2010) (2) 429-446.

[28] M. A. Santos da Silva, R. J. S. Maciel, L. N. Matos and M. H. Galina, TerraSOM: Sistema para Análise de Dados Geoespaciais Agregados por Área Baseado na Rede Neural do Tipo Mapa Auto-Organizável de Kohonen, Embrapa Tabuleiros Costeiros, Aracaju, 2015, p. 38.

[29] M. Halkidi and M. Vazirgiannis, A density-based cluster validity approach using multi-representatives, *Pattern Recognit. Lett.* 29 (2008) 773-786.

[30] S. Wu and T. W. Chow, Clustering of the self-organizing map using a clustering validity index based on inter-cluster and intra-cluster density, *Pattern Recognit.* 37 (2004) (2) 175-188.

[31] M. Cottrell and P. Letrémy, How to use the kohonen algorithm to simultaneously analyze individuals and modalities in a survey, *Neurocomputing* 63 (2005) 193-207.

[32] M. Cottrell, S. Ibbou and P. Letrémy, Som-based algorithms for qualitative variables, *Neural Networks* 17, (2004) (8) 1149-1167.

[33] F. Lourenço, V. Lobo and F. Bação, Binary-based similarity measures for categorical data and their application in self-organizing maps, in: *Jornadas de classificação e análise de dados*, 2004, p. 11.

[34] E. López-Rubio, Probabilistic self-organizing maps for qualitative data, *Neural Networks* 23 (2010) (10) 1208-1225.

[35] C. C. Hsu, S. H. Lin and W. S. Tai, Apply extended self-organizing map to cluster and classify mixed-type data, *Neurocomputing* 74 (2011) (18) 3832-3842.

[36] L. Jiao, Y. Liu, and B. Zou, Self-organizing dual clustering considering spatial analysis and hybrid distance measures, *Sci. China Earth Sci.* 54 (2011) (8) 1268-1278.

[37] M. S. Yang, W. L. Hung, and D. H. Chen, Self-organizing map for symbolic data, *Fuzzy Sets Syst.* 203 (2012) 49-73.

[38] C. de. Coso, D. Fustes, C. Dafonte, F. J. Nóvoa, J. M. Rodríguez-Pedreira, and B. Arcay, Mixing numerical and categorical data in a self-organizing map bymeans of frequency neurons, *Appl. Soft Comput.* 36 (2015) 246-254.

[39] A. Ultsch, *Self-Organizing Neural Networks for Visualisation and Classification*, Springer, 1993.

[40] R. Henriques, F. Bação and V. Lobo, Exploratory geospatial data analysis using the GeoSOM suite, *Comput. Environ. Urban Syst.* 36 (2012) 218-232.

[41] M. J. Greenacre, *Theory and Applications of Correspondence Analysis*, London: Academic Press, 1984.

[42] J. P. Benzécri, *Correspondence Analysis Handbook*, New York: Dekker, 1992.