

# Computational Methods for Agricultural Research: Advances and Applications

Hercules Antonio do Prado

*Brazilian Agricultural Research Corporation & Catholic University of  
Brasilia, Brazil*

Alfredo Jose Barreto Luiz

*Brazilian Agricultural Research Corporation, Brazil*

Homero Chaib Filho

*Brazilian Agricultural Research Corporation, Brazil*

Information Science  
**REFERENCE**

**INFORMATION SCIENCE REFERENCE**

Hershey • New York

Director of Editorial Content: Kristin Klinger  
Director of Book Publications: Julia Mosemann  
Acquisitions Editor: Lindsay Johnston  
Development Editor: Joel Gamon  
Typesetter: Deanna Jo Zombro  
Production Editor: Jamie Snavelly  
Cover Design: Lisa Tosheff

Published in the United States of America by  
Information Science Reference (an imprint of IGI Global)  
701 E. Chocolate Avenue  
Hershey PA 17033  
Tel: 717-533-8845  
Fax: 717-533-8661  
E-mail: [cust@igi-global.com](mailto:cust@igi-global.com)  
Web site: <http://www.igi-global.com>

Copyright © 2011 by IGI Global. All rights reserved. No part of this publication may be reproduced, stored or distributed in any form or by any means, electronic or mechanical, including photocopying, without written permission from the publisher. Product or company names used in this set are for identification purposes only. Inclusion of the names of the products or companies does not indicate a claim of ownership by IGI Global of the trademark or registered trademark.

#### Library of Congress Cataloging-in-Publication Data

Computational methods for agricultural research : advances and applications /  
Hercules Antonio do Prado, Alfredo Jose Barreto Luiz, and Homero Chaib Filho,  
editors.

p. cm.

Includes bibliographical references and index.

ISBN 978-1-61692-871-1 (hardcover) -- ISBN 978-1-61692-873-5 (ebook) 1.  
Agriculture--Research--Data processing. 2. Agricultural informatics. I.  
Prado, Hercules Antonio do. II. Luiz, Alfredo Jose Barreto, 1963- III. Filho,  
Homero Chaib.

S540.D38C66 2010

630.72--dc22

2010035367

#### British Cataloguing in Publication Data

A Cataloguing in Publication record for this book is available from the British Library.

All work contributed to this book is new, previously-unpublished material. The views expressed in this book are those of the authors, but not necessarily of the publisher.

## Chapter 7

# Using Self-Organizing Maps for Rural Territorial Typology

**Marcos Aurélio Santos da Silva**

*Brazilian Agricultural Research Corporation, Embrapa Coastal Tablelands, Brazil*

**Edmar Ramos de Siqueira**

*Brazilian Agricultural Research Corporation, Embrapa Coastal Tablelands, Brazil*

**Olívio Alberto Teixeira**

*Federal University of Sergipe, Department of Economy, Cidade Universitária, Brazil*

**Maria Geovania Lima Manos**

*Brazilian Agricultural Research Corporation, Embrapa Coastal Tablelands, Brazil*

**Antônio Miguel Vieira Monteiro**

*National Institute for Space Research, Image Processing Division, Brazil*

### ABSTRACT

*This work assessed the capacity of the self-organizing map, an unsupervised artificial neural network, to aid the process of territorial design through visualization and clustering methods applied to a multivariate geospatial temporal dataset. The method was applied in the case study of Sergipe's institutional regional partition (Territories of Identity). Results have shown that the proposed method can improve the exploratory spatial-temporal analysis capacity of policy makers that are interested in territorial typology. A new partition for rural planning was elaborated and confirmed the coherence of the Territories of Identity.*

### INTRODUCTION

Rural territorial typology refers to the classification of commonalities found in contiguous rural areas that may be useful for a well definition of

public policies and to discover identity elements. In an integrational perspective, all dimensions must be considered in a spatial analysis. For this study, it was preferred to adopt the definition of territory used by the Ministry of Agrarian Development (MDA, 2005, p.7).

DOI: 10.4018/978-1-61692-871-1.ch007

*... territories are defined as a physical space, geographically defined, generally continuous, encompassing urban and rural, characterized by multidimensional criteria, such as environment, economy, society, culture, politics and institutions, and a population with relatively distinct social groups that relate internally and externally through specific processes, which can distinguish one or more elements that indicate identity and social, cultural and territorial cohesion.*

Territories break down false boundaries (political divisions) among areas and facilitate the process of socio-cultural identity construction, discovery or investigation. The design of public policies also benefits from this strategy, in a way that watersheds, forests, ecosystems and other homogeneous zones can be analyzed altogether. Naturally, there are convergences and divergences about territorial definition, mainly concerning the focus of the study, for example, economical versus environmental analysis. A regional development study implies the use of an interdisciplinary approach, which reveals itself as a very difficult task due to the complexity and uncertainty of this field. Regional studies comprehend constraining factors, such as scales, physical amplitude, quantity and quality of data, diversity of applications, political and practical aspects (Sabourin & Teixeira, 2002).

The general definition of territory by the MDA (2005) argues that social-cultural-economical cohesion is the main factor to establish a regional aggregation into a territory. Although landscape homogeneity and distribution of natural resources are important aspects, they are not determinant. Therefore, multivariate socio-economical geospatial data integration and interpretation can be useful for policy makers that engage themselves in the process of territory design.

In Brazil, there are many initiatives towards the institutionalization of territories as a new strategy for public management (MDA, 2005; Bandeira, 2006). In this context of interdisciplinary studies and spatial integration, mainly for sustainable de-

velopment, rural and urban issues are not viewed as different things, but treated as a systematic problem (Sabourin & Teixeira, 2002; Flores, 2004; MDA, 2005). Besides, the new trend for self-development by bottom-up as stated by Claval (2008) increases the demand for new and practical approaches to study and generate knowledge for a good territorial partition.

The absence of a territorial theory (Abramovay, 2006) and the rise of territorial policies, at least in Brazil, imply the demand for exploratory methods that could solve practical problems, such as exploratory spatial clustering and, at the same time, it gives us helpful insights for a territorial theory. There are many classical quantitative strategies to perform a multivariate exploratory analysis of geospatial data such as: colored maps, multivariate statistics coupled with these maps, spatial statistics, and geostatistics (Bailey, 1995). In this study, an Artificial Neural Network (ANN) was applied simultaneously for geospatial data visualization and clustering. The objective was to assess the capacity of the ANN, more precisely the Self-Organizing Map (SOM), to find hidden patterns in the geospatial dataset that could be useful for a territorial typology.

This ANN was applied in the analysis of Sergipe's institutional regional partition (Territories of Identity) created by Teixeira et al. (2007) using both subjective and analytical methods. On the other hand, the approach of this work used quantitative geospatial dataset aggregated by municipalities. Results were compared with this official Sergipe's clustering. These territories were chosen because a great effort for Sergipe's territorial partition was performed by academics and officials, so a huge amount of information was gathered and organized. Consequently, the know-how and knowledge about the way the study of regional commonalities is performed increased and, therefore, demanded for integrated quantitative analysis of multisource and multivariate geospatial data. In general terms, the main goal is to establish a semi-automatic territorial typology

process to aid the rural planning by identification of homogeneous areas in Sergipe, Brazil.

## MATERIAL AND METHODS

### Self-Organizing Maps (SOM)

The Kohonen Self-Organizing Map (SOM) is a competitive artificial neural network structured in two layers (Kohonen, 2001). The first one represents the input vector of the dataset variables  $x_k = [\zeta_1, \zeta_2, \dots, \zeta_d]^T$ ,  $k=1, 2, \dots, n$ , where  $n$  is the number of input vectors and  $d$  corresponds to data dimension, and the second one is a neuron's grid, usually bidimensional and fully connected (Figure 1a). Each neuron  $j$  has one codevector  $w_j = [w_{j1}, w_{j2}, \dots, w_{jd}]^T$  associated to it. The main goal of this ANN is to order, by a learning algorithm, the input dataset into the neural grid, using the codevectors ( $w_j$ ) as representations of the input data ( $x_k$ ).

There are different topologies for the Self-Organizing Map architecture, while the most common is the two-dimensional topology. Figure 1 shows a one-dimensional SOM (b) and a toroidal SOM (c).

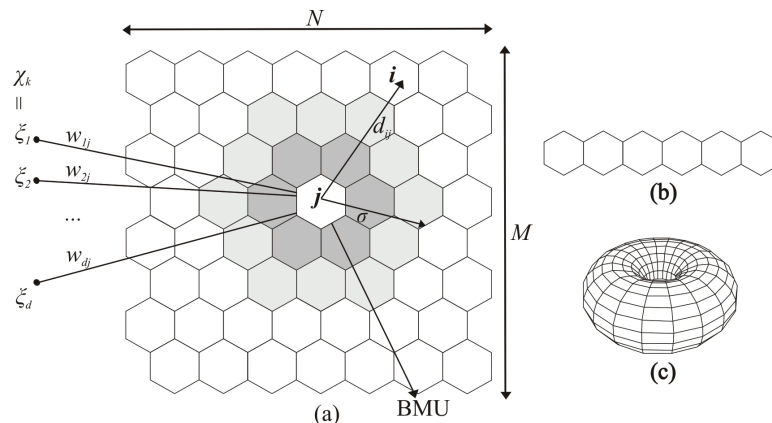
The learning process can be explained in three phases. In the first phase (competitive) each  $x_k$  is presented to all neurons in search of the Best Match Unit (BMU) using Euclidean distance as a measurement reference of the distance feature. In the second phase (cooperative) a neighborhood relation between the BMU and the other neurons is defined for a smoothed update of each  $w_j$ . Finally, in the last phase (adaptive), all codevectors  $w_j$  will be updated using some kind of adaptive rule, in general sequential or batch, see Equations 1 and 2 for the batch update rule used in this work (Vesanto, 1999).

$$s_i(t) = \sum_j^{n_{V_i}} x_i \quad (1)$$

$$w_i(t+1) = \frac{\sum_j^m h_{ji}(t) s_j(t)}{\sum_j^m n_{V_j} h_{ji}(t)} \quad (2)$$

Where  $s_i$  represents an input pattern sum for the  $i^{th}$ -Voronoi region,  $V_i$ , and  $n_{V_i}$  is the number of samples for the Voronoi dataset of the  $i^{th}$ -neuron. The neighborhood kernel function between neu-

Figure 1. (a) A complete self-organizing  $N \times M$  map architecture representation. (b) A one-dimensional SOM. (c) A toroidal SOM



rons  $i$  and  $j$  at time  $t$  is represented by  $h_{ij}(t) = \exp(-d_{ij}^2/2 * \delta(t)^2)$ , where  $\delta(t) = \delta(0) * \exp(-t/C)$ ,  $d_{ij}$  is the distance between neurons  $i$  and  $j$  on the neural grid,  $C$  is a constant, and  $m$  is the number of Voronoi regions (number of neurons). See Equation 3 for the sequential update rule.

$$w_{ji}(t+1) = w_{ji}(t) + \alpha(t) h_{ij}(t) [x_{ik}(t) - w_{ij}(t)] \quad (3)$$

Where  $\alpha(t)$ ,  $\alpha(0) < 1$ , represents the learning rate function. The learning rate function  $\alpha(t)$ ,  $h_{ij}(t)$  and  $\delta(t)$  are monotonically decreasing functions.

After the learning process, codevectors should approximate, in a non-linear manner, the input dataset. In addition, SOM preserves the topological structure of the input data, and nearby patterns in the sample dataset are associated with nearby neurons in the SOM grid. For this work, a bi-dimensional SOM, the batch learning algorithm associated with hexagonal lattice, gaussian neighborhood kernel function, and linear initialization for visualization were used. For the automatic clustering it was used a uni-dimensional SOM.

The competitive process is the most time-consuming of the learning process. Usually this is a sequential search for the Best Match Unit (BMU). This process can be optimized by using some mechanism to minimize the heuristic search or through the parallelization of the learning code (Openshaw & Turton, 1996).

The parameters of learning are defined empirically, based on user experience and methods of trial and error. However, some techniques for automatically determining the parameters of learning have been proposed, either through genetic algorithms or numerical methods (Haese & Goodhill, 2001).

The dimensionality of self-organizing map and its size ( $m$ ) depend on the type of problem and purpose. The literature shows that determining the size of the SOM is an empirical process (Flexer, 2001; Kohonen, 2001). In general, bi-dimensional SOM is used because of its ability to project the

data of high dimension  $d$  into a two-dimensional neural map grid.

## Assessment of Quality of SOM Learning Algorithm

There is a reasonable set of methods for assessing the quality of the generated map after the learning process. The most applied are the quantization error vector and the topological error (Kohonen, 2001).

The quantization error ( $E_q$ ), Equation 4, is the average error of values of the difference between the feature vector  $x_k$  and  $w_{BMU}$  codevector, that is the winner in the competitive process for the vector  $x_k$ :

$$E_q = \frac{\sum_{k=1}^n \|x_k - w_{BMU}\|}{n} \quad (4)$$

The topological error ( $E_t$ ), Equation 5, evaluates how the structure of the neural grid approximate vectors of the input space. Whereas each  $x_k$  has a BMU as the first neuron in the order of competition in the neural grid, BMU2 correspond to second neuron at this scale. Thus, the error will correspond to the percentage of BMU and BMU2 whose are not neighbors in the grid for the same  $x_k$ :

$$E_t = \frac{1}{n} \sum_{k=1}^n u(x_k) \quad (5)$$

Where  $u(x_k)$  is setted to one if the BMU and BMU2 are not neighbors, and zero otherwise.

## Data Visualization

After the learning process it is necessary to visually verify the result of topological ordering onto the SOM's grid. There are many ways of visual representation of the dataset using SOM (Vesanto, 1997; Vesanto, 1999). For this study, two of the

most common methods were used. The first is the direct link between the artificial neuron unit and the  $x_k$ , in this case, the municipality.

The second form, the Component Planes, uses the values of each component of the codevector  $w_j$  to color the Self-Organizing Map. This method allows the visual analysis of component correlation, outlier detection and spatial component distribution on SOM (Vesanto, 1999; Kohonen, 2000; Silva et al., 2004a).

Two periods were used and such temporal aspect was included in the SOM model by mounting these two datasets in only one, so each temporal set of values for each municipality became an entry  $x_k$ . This is the simplest strategy, but yet useful when the purpose in mind is a visual clustering and interpretation by component planes. According to Barreto & Araújo (2001), there is a plethora of SOM models for time processing, but the most common application of the SOM in sequence processing is the trajectory of BMU's in the latent map space (Varsta, 2002). Therefore, these two periods were compared visually by Component Planes following the trajectory of each municipality onto the neural map. To proceed this visual analysis, the last version of the SOM ToolBox was used ([www.cis.hut.fi/somtoolbox/](http://www.cis.hut.fi/somtoolbox/)).

## Automatic Clustering

The SOM's codevectors can also be segmented by a wide range of strategies, including the visual ones (Vesanto & Alhoniemi, 2000). Costa (1999) partitioned the dataset throughout the segmentation of the image generated by the codevector's matrix of distance. However, in general, this is too complex and it does not exhibit clearly hidden patterns. Vesanto (1999) used SOM as a data compressor for a posterior statistical analysis, but this strategy needs a high level of human interaction. Costa & Andrade Netto (2003) used a conceptually simple and empiric approach (Costa-Netto algorithm) using only the information stored in codevectors after the learning process as the activ-

ity level of neurons, the quantization error and the neighborhood relation for an automatic clustering.

Considering the output layer as a non-oriented graph structure, Costa & Andrade Netto (2003) proposed a SOM clustering based on graph segmentation, in a way that the algorithm eliminates inconsistent connections among neurons, in attempt to find the ideal partition of the SOM's grid. The algorithm is based on the geometric distance information among neurons, quantization error and level of the neuron's activity. The strategy is to use heuristic rules to eliminate inconsistent connections between two neighbor neurons.

The algorithm:

- 1.1. Take all distances  $d$  between adjacent neurons  $i$  and  $j$  using the codevector  $w_j$  as a measurement parameter,  $d(w_i, w_j)$ ; and activity level for each neuron  $i$ ,  $H(i)$ .
- 1.2. For each couple of adjacent neurons,  $i$  and  $j$ , the edge will be considered inconsistent if:
  - 1.2.1. the distance between codevectors exceed by 2 the median distance of the other adjacent neurons to  $i$  or  $j$ ;
  - 1.2.2. two adjacent neurons have the level of activity ( $H$ ) less than 50% of the allowed minimum ( $H_{min}$ ), or some of them are inactive neurons ( $H(i) = 0$ );  $H_{min} = \omega H_{med}$ ,  $0.1 \leq \omega \leq 0.6$  and  $H_{med} = n/m$ ;
  - 1.2.3. the distance between the centroids of the dataset associated to neurons  $i$  and  $j$  exceeds by 2 the distance between the codevectors of  $i$  and  $j$ ,  $d(w_i, w_j)$ .
- 1.3. Remove all inconsistent edges.
- 1.4. Associate a unique code for each set of connected neurons.

At the end of the process, each group of connected neurons will represent one cluster of the codevectors. The algorithm uses some empirical values, defined by the user, but can segment the input data using only the SOM information after the

learning process. To validate the clusters partition of the dataset, it was used the Compose Density between and within the clusters (CDbw) index (Halkidi & Vazirgiannis, 2002; Silva et al., 2004b; Wu & Chow, 2004). For this spatial analysis, the TerraView software ([www.dpi.inpe.br/terraview](http://www.dpi.inpe.br/terraview)) was used, as well as the TerraSOM plugging developed from the SOMCode library ([www.cpatc.embrapa.br/somcode/](http://www.cpatc.embrapa.br/somcode/)) that implements some Self-Organizing Map related algorithms.

### Dataset

The State of Sergipe has an area of 21,910.34 km<sup>2</sup>, accounting for 0.26% of national territory and 1.4% of the Northeast. Its absolute position is between the parallel 9° 31' and 11° 34' south latitude and the meridians 36° 25' and 38 ° 14' longitude West of Greenwich (Figure 2). Although it constitutes a state of small size and population, Sergipe occupies a privileged position in the economic and social development scenario

of the Northeast, considering its recent entry into modern economic activities: off-shore oil exploitation, mining, tourism and ethanol production by sugar cane crops. The agriculture share in the GDP corresponds to approximately 3,0%, but it is very important in terms of employment and food supply.

Sergipe's economy is diversified. Nevertheless, it is important to highlight the mineral extractive-industry (based on non-metallic minerals) located in cities within a radius of up to 40 km from the capital, which intensely contributes to the concentration of the population, income and wealth in this area. Agricultural activities are distributed throughout the whole territory, occupying a significant area, as well as labor force, with a strong emphasis on family work. Sugar cane and orange stand out, as well as cattle and poultry. The tertiary sector has been showing changes with the addition of new functions, especially in the service sector (Teixeira et al., 2007).

Figure 2. Municipalities of Sergipe's state

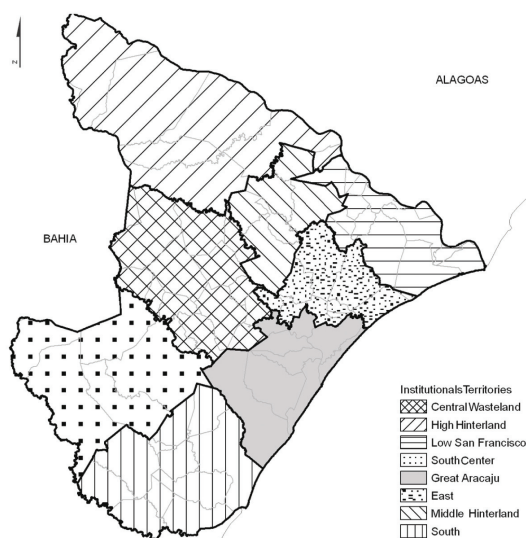




For this study, 44 indicators (Table 1) were selected to represent demographic, social, economic, and agricultural production dimensions, as well as agricultural activities, such as land use and crop and animal production. Two periods (A and B) were considered in this study for a comparison between the 1990's and 2000's values. All variables were standardized according to their meaning and standard deviation. These variables were chosen according to their disponibility and proximity with those used by Teixeira et al. (2007).

The typology process used the institutional partition of the State of Sergipe held by the Department of Planning as a starting point, named Territory of Identity. This process takes into account quantitative indicators (Table 1), popular opinions gathered by workshops and multidisciplinary assessment of spatial aggregation (Teixeira et al., 2007). The main goal of this partition was to find out a well-established division of the State that could aid the process of general public policies design. Figure 3 shows the eight Territories of Identity: Central Wasteland, High Hinterland, Low San Francisco, South Center, Great Aracaju, East, Middle Hinterland and South.

Figure 3. Territories of identity



The **High Hinterland Territory** is formed by seven municipalities: Canindé do São Francisco (CSF), Grararu (GRR), Monte Alegre de Sergipe (MAS), Nossa Senhora da Glória (GLR), Nossa Senhora de Lourdes (LOU), Poço Redondo (PRD) and Porto da Folha (PFL). This territory represents 22.3% of the states's land area, covering 4908 km<sup>2</sup>. In 2007, the High Hinterland Territory had a population of 137,926 inhabitants, with a density of 28.15 hab/Km<sup>2</sup>, representing 7.11% of the total population of the state. In 2005, the GDP of this Territory represented 11.12% of the state's GDP, with a GDP per capita of R\$ 10,730.00. The Index of Human Development in 2000 was 0.575.

The **Low San Francisco Territory** is formed by fourteen municipalities: Amparo do São Francisco (ASF), Brejo Grande (BRJ), Canhoba (CNH), Cedro de São João (CED), Ilha das Flores (IFL), Japoatã (JPT), Malhada dos Bois (MLH), Muribeca (MUR), Neópolis (NEO), Pacatuba (PAC), Propriá (PRO), Santana do São Francisco (SSF), São Francisco (SFR) and Telha (TEL). This territory corresponds to 9.0% of the state's land area, which represents 1986.3 km<sup>2</sup> of total area. In 2007, the Low San Francisco Territory had a population of 123,482 inhabitants, with a density of 63.45 hab/Km<sup>2</sup>, representing 6.37% of the population of the state. In 2005, the GDP of this Territory represented 3.94% of the state's GDP, with a GDP per capita of R\$ 4,176.00. The Index of Human Development in 2000 was 0.614.

The **Middle Hinterland Territory** is formed by six municipalities: Aquidabã (AQB), Cumbe (CUM), Feira Nova (FNV), Graccho Cardoso (GCR), Itabi (ITB) and Nossa Senhora das Dores (DOR). This territory represents 7.3% of the land area of the state, corresponding to 1612.6 km<sup>2</sup> of total area. In 2007, the Middle Hinterland Territory had a population of 62,644 inhabitants, with a density of 39.59 hab/Km<sup>2</sup>, representing 3.23% of the population of the state. In 2005, the GDP of this Territory represented 1.57% of the state's GDP, with a GDP per capita of R\$ 3,298.00. The Index of Human Development in 2000 was 0.621.

Table 1. List of analyzed variables

	VARIABLE	DESCRIPTION	GROUP A	GROUP B
1	POPTOT	Total of resident population (people)	Pop. 1991 <sup>1</sup>	Pop. 2000 <sup>2</sup>
2	POPRUR	Total of rural resident population (people)	Pop. 1991	Pop. 2000
3	PRODMCOW	Milk cow production (1.000 l)	Census 1996 <sup>3</sup>	Census 2006 <sup>4</sup>
4	PRODMGOAT	Milk goat production (1.000 l)	Census 1996	Census 2006
5	PRODEGGS	Egg production (t)	Census 1996	Census 2006
6	TOTCOX	Cattle - Ox (unit)	Census 1996	Census 2006
7	TOTCS	Cattle - swine (unit)	Census 1996	Census 2006
8	TOTBIRDS	Birds (unit)	Census 1996	Census 2006
9	TRACTORS	Tractors (amount)	Census 1996	Census 2006
10	PAREAPERM	Permanent and temporary crops (ha)	Census 1996	Census 2006
11	PAREAPAST	Natural and artificial pastures (ha)	Census 1996	Census 2006
12	PAREAFOT	Cultivated and natural forests (ha)	Census 1996	Census 2006
13	PMCFJSJ	Percentage of women heads of households with no spouse and with minor children (15 years old)	ATLAS 1991 <sup>5</sup>	ATLAS 2000 <sup>5</sup>
14	PPOOR	Percentage of poor people	ATLAS 1991	ATLAS 2000
15	PCI	Percentage of indigent children	ATLAS 1991	ATLAS 2000
16	PC714FE	Percentage of children out of school (range from 7 to 14 years old)	ATLAS 1991	ATLAS 2000
17	INPERCAPITA	Per capita income (R\$)	ATLAS 1991	ATLAS 2000
18	PINGOV	Income's percentage related to governmental transferences	ATLAS 1991	ATLAS 2000
19	PINWORK	Income's percentage related to work's yield	ATLAS 1991	ATLAS 2000
20	RBRICSPOOR	Relation between 10% richest and 40% poorest	ATLAS 1991	ATLAS 2000
21	GINI	Gini's index	ATLAS 1991	ATLAS 2000
22	ITHEIL	Theil's index	ATLAS 1991	ATLAS 2000
23	INTIND	Indigence intensity	ATLAS 1991	ATLAS 2000
24	IDHMIN	Municipal Human Development Index - Income	ATLAS 1991	ATLAS 2000
25	IDHMLONG	Municipal Human Development Index - Longevity	ATLAS 1991	ATLAS 2000
26	ESPLIFE	Life expectancy at birth	ATLAS 1991	ATLAS 2000
27	IDHMEDUC	Municipal Human Development Index - Education	ATLAS 1991	ATLAS 2000
28	RILLI	Rate of illiteracy	ATLAS 1991	ATLAS 2000
29	GRSA	Gross rate of school attendance	ATLAS 1991	ATLAS 2000
30	RFECUN	Rate of fecundity	ATLAS 1991	ATLAS 2000
31	PPDOWN	Percentage of people living in their own and paid off house or land	ATLAS 1991	ATLAS 2000
32	P714A	Percentage of illiterate children (ranging from 7 to 14 years old)	ATLAS 1991	ATLAS 2000
33	PPWATER	Percentage of people living in houses with piped water	ATLAS 1991	ATLAS 2000
34	PPWC	Percentage of people living in houses with WC and piped water	ATLAS 1991	ATLAS 2000
35	PPGARB	Percentage of people living in houses with garbage service	ATLAS 1991	ATLAS 2000
36	PPENERGY	Percentage of people living in houses with electric energy	ATLAS 1991	ATLAS 2000
37	PPIBAGRI	Percentage of Agricultural GDP	IPEA 1996 <sup>6</sup>	IPEA 2006 <sup>6</sup>

*continued on following page*

Table 1. continued

	VARIABLE	DESCRIPTION	GROUP A	GROUP B
38	RTTotal	Constitutional transferences from National Treasure	MF 1996 <sup>7</sup>	MF 2006 <sup>7</sup>
39	MPRODMAIZE	Maize production (t)	Mean 90/96 <sup>8</sup>	Mean 00/06 <sup>8</sup>
40	MPRODCAS	Cassava production (t)	Mean 90/96	Mean 00/06
41	MPRODBean	Bean production (t)	Mean 90/96	Mean 00/06
42	MPRODSC	Sugar cane production (t)	Mean 90/96	Mean 00/06
43	MPRODORAN	Orange production (t)	Mean 90/96	Mean 00/06
44	MPRODCOCO	Coconut production (1.000 fruits)	Mean 90/96	Mean 00/06

SOURCE: <sup>1</sup>Demographic census 1991 - IBGE. <sup>2</sup>Demographic census 2000 - IBGE. <sup>3</sup>Agricultural census 1995/1996 - IBGE. <sup>4</sup>Agricultural census 2006 - IBGE. <sup>5</sup>Atlas of Human Development - PNUD/Brasil. <sup>6</sup>IPEDATA (<http://www.ipeadata.gov.br/>). <sup>7</sup>Ministry of Finance - National Treasure ([http://www.tesouro.fazenda.gov.br/estados\\_municipios/transferencias\\_constitucionais.asp](http://www.tesouro.fazenda.gov.br/estados_municipios/transferencias_constitucionais.asp)). <sup>8</sup>Anual Municipal Agricultural Production - PMA/IBGE.

The **East Territory** is formed by nine municipalities: Capela (CPL), Carmópolis (CAR), Divina Pastora (DVN), General Maynard (GMN), Japaratinga (JTB), Pirambu (PIR), Rosário do Catete (RCT), Santa Rosa de Lima (SRL) and Siriri (SRR). This territory corresponds to 6.7% of the land area of the state, possessing 1474.1 km<sup>2</sup> of total area. In 2007, the territory of East Sergipe had a population of 90,452 inhabitants, with a density of 59.56 hab/Km<sup>2</sup>, representing 4.66% of the population of the state. In 2005, the GDP of this Territory represented 6.47% of the state's GDP, with a GDP per capita of R\$ 9,878.00. The Index of Human Development in 2000 was 0.643.

The **Central Wasteland Territory** is formed by fourteen municipalities: Areia Branca (ARB), Campo do Brito (CPB), Carira (CRR), Frei Paulo (FPA), Itabaiana (ITA), Macambira (MAC), Malhador (MLD), Moita Bonita (MOI), Nossa Senhora Aparecida (APA), Pedra Mole (PRM), Pinhão (PIN), Ribeirópolis (RIB), São Domingos (SDO) and São Miguel do Aleixo (SMA). This territory corresponds to 6.7% of the land area of the state, possessing 1474.1 km<sup>2</sup> of total area. In 2007, the Central Wasteland Territory had a population of 222,197 inhabitants, with a density of 71.14 hab/Km<sup>2</sup>, representing 11.46% of the population of the state. In 2005, the GDP of this Territory represented 6.70% of the state's

GDP, with a GDP per capita of R\$ 3,949.00. The Index of Human Development in 2000 was 0.646.

The **South Territory** is formed by eleven municipalities: Arauá (ARA), Boquim (BOQ), Cristinápolis (CRI), Estância (EST), Indiaroba (IND), Itabaianinha (ITN), Pedrinhas (PDR), Salgado (SAL), Santa Luzia do Itanhyl (SLI), Tomar do Geru (TGE) and Umbaúba (UMB). This territory represents 14.5% of the land area of the state, with 3193.6 km<sup>2</sup> of total area. In 2007, the South Territory had a population of 241,192 inhabitants, with a density of 77.06 hab/Km<sup>2</sup>, representing 12.44% of the population of the state. In 2005, the GDP of this Territory represented 8.40% of the state's GDP, with a GDP per capita of R\$ 4,579.00. The Index of Human Development in 2000 was 0.616.

The **South Center Territory** is formed by five municipalities: Lagarto (LGT), Poço Verde (PVD), Riachão do Dantas (RDT), Simão Dias (SDI) and Tobias Barreto (TBA). This territory represents 16.1% of the land area of the state, possessing 3551.5 km<sup>2</sup> of total area. In 2007, the South Center Territory had a population of 213,492 inhabitants, with a density of 60.63 hab/Km<sup>2</sup>, representing 16.07% of the population of the state. In 2005, the GDP of this Territory represented 5.67% of the state's GDP, with a GDP per capita of R\$ 3,480.00. The Index of Human Development in 2000 was 0.599.

The **Great Aracaju Territory** is formed by nine municipalities: Aracaju (AJU), Barra dos Coqueiros (BCQ), Itaporanga d'Ajuda (IAJ), Laranjeiras (LRJ), Maruim (MRI), Nossa Senhora do Socorro (SOC), Riachuelo (RIA), Santo Amaro das Brotas (SAB) and São Cristóvão (SCR). This territory represents 9.9% of the land area of the state, with 2192 km<sup>2</sup> of total area. In 2007, the Great Aracaju Territory had a population of 847,941 inhabitants, with a density of 387.65 hab/Km<sup>2</sup>, representing 43.72% of the population of the state. In 2005, the GDP of this Territory represented 56.12% of the state's GDP, with a GDP per capita of R\$ 8,785.00. The Index of Human Development in 2000 was 0.750.

## **GEOSPATIAL ANALYSIS BY SELF-ORGANIZING MAPS**

The analysis of multivariate geospatial data could be pointed as one of the great scientific challenges to be solved in the present. Several institutions are providing amounts of geo-referenced data over the Internet, but there are many limitations to the effective use and interpretation of that type of dataset. In fact, the integration of all information depends on the ability to analyze the data and find relevant knowledge in such databases. As stated by Bação et al. (2005) "More than prediction tools, we need to develop exploratory tools based on classification and clustering, which would enable an improved understanding of the available data" (p. 155).

The Self-Organizing Map has been applied to the problem of geospatial analysis in many ways. Nevertheless, in general, the SOM is used for exploratory spatial analysis of census data aggregated by area. The way the analysis is performed varies, but the use of visual interpretation methods is registered in almost every application.

Winter & Hewitson (1994) applied the SOM associated with the Component Planes for the characterization of 21 districts of the Western Cape

of South Africa using 10 census variables. From the Component Planes it was possible to characterize mainly the black population of the region. Kaski & Kohonen (1996) showed the feasibility of the visualization of multivariate data from the SOM. The dataset consisted of 39 indicators, which described different aspects of the welfare "states" of the countries around the world.

Koua (2003) analyzed 29 socioeconomic indicators related to municipalities. The author used the Component Planes and the matrix of distances among codevectors  $w_j$  for a visual geospatial exploration and discovered that SOM improved the geographical analysis and offered support for the exploration of large geospatial datasets, increasing knowledge in this area. Silva (2004ab) performed a complete exploratory spatial analysis using eight socio-economic indicators related to census sectors. The author adjusted the algorithms related to the SOM neural network in a suite for outlier detection, spatial dependence analysis, automatic clustering and component analysis. The suite was applied in the problem of intra-urban analysis of social exclusion and was compared with statistical analysis (Genovez, 2002).

Aksoy (2006) applied the SOM to the problem of the regionalization of 923 districts in Turkey using 36 socio-economic indicators and compared the results with a statistical approach (k-means and Principal Component Analysis). Spielman & Thill (2008) studied the spatial relationship among 2,217 census tracts in New York City using a dataset of 79 attributes applying a simple method of geospatial data visualization linking the artificial neural network with a Geographical Information System (GIS).

There are some differences between the approach of this chapter and the others listed in this short review. First, in the present study, there is an explicit focus on rural aspects of territorial study while the majority of researches focus on urban problems; second, the chapter proposed a complete regional characterization (typology) using, at the same time, visual tools and an automatic clustering

algorithm; finally, the work started from a previous qualitative territorial study, Territories of Identity, that makes a comparison of the results possible.

## RESULTS AND DISCUSSION

### General Spatial Dependence by Visual Interpretation

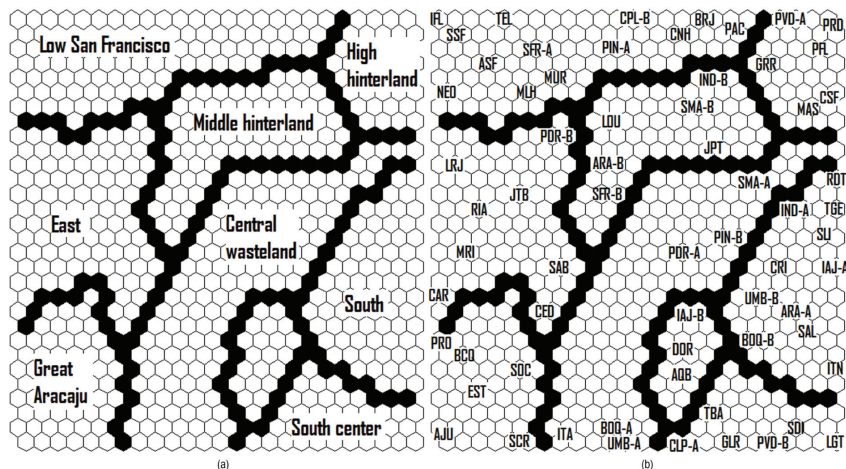
After SOM's learning algorithm processing had been performed, the correspondence between each input vector (municipality-group) to its corresponding artificial neuron was realized, according to the Euclidean distance. Thus, it was possible to visualize the position in the neural map of each municipality-group and examine the proximity relations in the feature's space among them. Whereas the study has the institutional division of the state of Sergipe into Territories of Identity as its starting point, the neural map was partitioned having it as guide, which means looking for proximities among municipalities of the same Territory of Identity.

From the visual analysis of the distribution of the cities on the neural map it was found that

there is spatial dependence for most of the input vectors. That is, neighboring municipalities in geophysical space are also deeply related in the space of attributes (neural map). From this observation it was possible to trace thresholds among the regions of the neural map that corresponded to most municipalities-group of each area previously defined (Territories of Identity). Therefore, Figure 4a illustrates how the Territories of Identity are distributed in the neural map. It was also observed that there is spatial dependence among these territories.

Figure 4b shows the distribution of some municipalities in the neural map so that it is possible to analyze two factors: discrepancies between the positions of each municipality according to data from two periods (A and B), and the interterritorial shifting of municipalities in the space of attributes, so that it is possible to analyze the link between the territories constructed by the neural map and the Territories of Identity. It is important to note that when significant differences between the groups A and B for the same city are not present, it will be only represented by the symbol of the city. Otherwise, the municipal-

Figure 4. (a) SOM after the visual partition. (b) Linking between each artificial neuron and the municipality associated to it. The number of municipalities represented onto the neural map was limited for legibility reasons



ity will be represented by its initials followed by a letter indicating the group (A or B).

From Figure 4b, it is possible to identify the local displacement between periods A and B. The municipality of Capela (CPL) migrated from the region linked to the South Central region to the Lower San Francisco Territory, approaching the municipalities of the East. The municipality of Poço Verde (PVD) migrated from the area linked to the High Hinterland to the area linked to the South Center. The municipality of São Miguel do Aleixo (SMA) migrated from the region associated with the Central Wasteland to the Middle Hinterland. The municipalities of Umbaúba (UMB) and Boquim (BOQ) migrated from the Central Wasteland to the South region. The municipality of Itaporanga d'Ajuda (IAJ) migrated from the South region to the region nearby the Central Wasteland, into the direction of the Great Aracaju. The municipalities of Indiaroba (IND), Pedrinhas (PDR) and Arauá (ARA) scattered from the area linked to the South, contrary to the behavior of BOQ and UMB.

The analysis of temporal shifts between groups A and B did not show significant changes for most municipalities. The Southern region showed greater variation between the two periods in the sense of closeness between municipalities (BOQ and UMB) and distance (IND, PDR, NIV). The municipalities of CPL and PVD had the highest displacements. The municipality of São Francisco (SFR) migrated from the Low San Francisco to the region associated to the Middle Hinterland.

Only half the regions (that represent the territories) formed by the neural map have at least one of its Centralities<sup>1</sup> not positioned in the region of concentration of the majority of the municipalities of the same Territory of Identity. The municipality of Estancia (EST), for example, belongs to the Southern Territory but is extremely close to the Great Aracaju Territory on the neural map. The municipalities of Aquidabã (AQB) and Nossa Senhora das Dores (DOR) are far away in the feature space of the other municipalities

of the Middle Hinterland Territory. Canindé do São Francisco (CSF) is a centrality with characteristics similar to those of most municipalities in the High Hinterland Territory, which does not occur with the municipality of Nossa Senhora da Glória (GLR), that stands in the region of the South Central Territory. Another centrality that is positioned distantly from the other municipalities of the same Territory in the feature space is Propriá (PRO), which also has similar characteristics to the municipalities of the Great Aracaju. All the centralities are shown in Figure 3b.

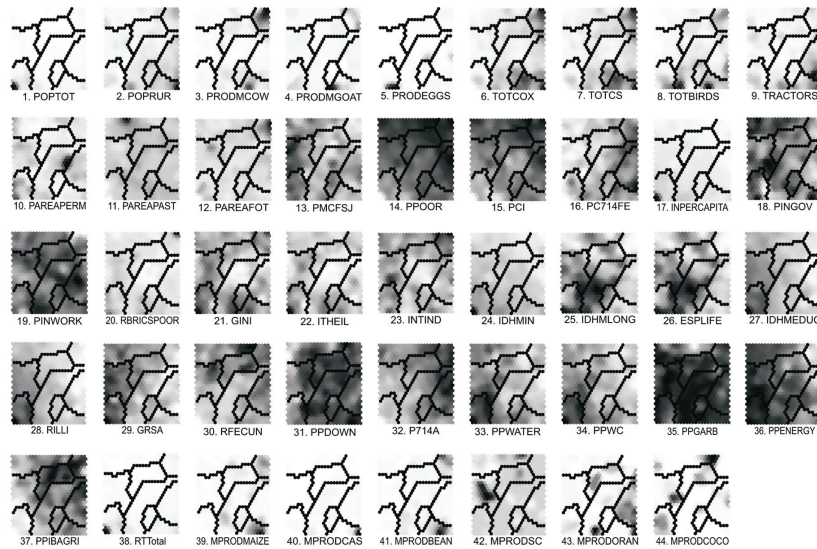
The same occurs with other municipalities that are not Centralities. The municipality of Riachão do Dantas (RDT) belongs to the South Central Territory, but has features related to the South Territory. The municipality of Nossa Senhora de Lourdes (LOU) belongs to the High Hinterland Territory, but has characteristics more intimately related to the Middle Hinterland Territory. The municipalities of Santo Amaro das Brotas (SAB), Maruim (MRI), Riachuelo (RIA) and Laranjeiras (LRJ) are associated to the Great Aracaju Territory, in spite of presenting characteristics associated with the municipalities of the East Territory.

It was observed that there are some differences between the groups formed by the neural map and the Territories of Identity, although, in general, the neural approach confirms the cohesion intra-Territories in the feature space.

### **Characterization by Component Planes**

Through Components Planes (Figure 5) it was possible to characterize each group defined by neural map (that represents each territory) through the observation of the spatial distribution of each component intensity in the neural map. As each municipality is associated with only one artificial neuron it is possible to use the partition of the neural map to facilitate the characterization of the territories.

Figure 5. Component planes for each variable



First, it was observed that there were outliers for some variables. Aracaju (AJU) was differentiated from the others with high values for the variables POPTOT, INPERCAPITA, ITHEIL, IDHMIN and RTTOTAL. The municipality of Lagarto (LGT) has high scores for the variables POPRUR and MPRODCCAS. The High Hinterland and South Center territories presents high values for PRODMCOW, PRODMGOAT, TOTCOX, TOTBIRDS, TOTCS, MPRODMAIZE and MPRODBEAN. Outliers were also detected for PRODEGGS, PAREAPAST, PAREAFOT, RBRICSPOOR variables. Correlation was also observed, same patterns for the Component Plane coloring, between variables PPOOR and PCI, IDHMLONG and ESPLIFE, IDHMEDUC and RILLI and PPWATER and PPWC.

The area which borders the High Hinterland region on the neural map was characterized by high values for the variables TOTCOX, PPOOR, PCI, INTIND, PPDOWN, P714A, PPGARB, PPIBAGRI and MPRODSC, average values for the variables PC714FE and GINI, and low values for the variables PAREAPERM, PMCFSJ, PINGOV, IDHMLONG, ESPLIFE, IDHMEDUC,

RILLI, GRSA, PPWATER, PPWC, PPENERGY, MPRODORAN and MPRODCOCO.

The area which borders the Lower San Francisco region was characterized by high values for the variables PPOOR, PCI, PINGOV, RFECUN and PPENERGY, average values for the variables PMCFSJ, PINWORK, GINI, INTIND, IDHMEDUC, RILLI, GRSA, PPDOWN, P714A, PPWATER, PPWC and PPIBAGRI, and low values for the variables TOTCOX, TOTBIRDS, TRACTORS, PAREAPERM, PC714FE, IDHMLONG, ESPLIFE, PPGARB, MPRODSC, MPRODORAN and MPRODCOCO.

The area which borders the Middle hinterland region was characterized by high values for the variables PPOOR, PCI, PINWORK, PPDOWN, PPGARB and PPIBAGRI, average values for the variables PINGOV, GRSA, RFECUN, P714A and PPENERGY, and low values for the variables TOTCOX, TRACTORS, PAREAPERM, PMCFSJ, PC714FE, GINI, INTIND, IDHMLONG, ESPLIFE, IDHMEDUC, RILLI, PPWATER, PPWC, MPRODSC, MPRODORAN and MPRODCOCO.

The area which borders the East region on neural map was characterized by high values for

the variables PMCFSJ, PPOOR, PCI, PPGARB, PPENERGY and MPRODSC, average values for the variables IDHMLONG, ESPLIFE, IDHMEDUC, RILLI, GRSA, RFECUN, PPDOWN, PPWATER and PPWC and low values for the variables TOTCOX, TRACTORS, PAREAPERM, PC714FE, P714A, PPIBAGRI, MPRODORAN and MPRODCOCO.

The area which borders the Central Wasteland region was characterized by high values for the variables IDHMLONG, ESPLIFE, PPDOWN, PPGARB and PPENERGY, average values for the variables PPOOR, PCI, PC714FE, PINGOV, PINWORK and PPIBAGRI, and low values for the variables TOTCOX, TRACTORS, PAREAPERM, PMCFSJ, GINI, INTIND, IDHMEDUC, RILLI, GRSA, RFECUN, P714A, PPWATER, PPWC, MPRODORAN and MPRODCOCO.

The area which borders the South region on the neural map was characterized by high values for the variables PPOOR, PCI, PC714FE, PINWORK, PPDOWN, P714A, PPGARB, PPIBAGRI, MPRODORAN, average values for the variables TRACTORS, PMCFSJ, GINI, RFECUN, and low values for the variables TOTCOX, PAREAPERM, PINGOV, INTIND, IDHMLONG, ESPLIFE, IDHMEDUC, RILLI, GRSA, PPWATER, PPWC, PPENERGY, and MPRODCOCO.

The area which borders the South Central region was characterized by high values for the variables TOTCOX, PPDOWN, and PPGARB, average values for the variables TRACTORS, PPOOR, PCI, PC714FE, PINGOV, PINWORK, GINI, INTIND, RFECUN, P714A, PPWATER, PPWC, PPENERGY and PPIBAGRI, and low values for the variables PAREAPERM, IDHMLONG, ESPLIFE, IDHMEDUC, RILLI, GRSA, MPRODORAN and MPRODCOCO.

The area which borders the Great Aracaju region on neural map is characterized by high values for the variables PINGOV, GINI, IDHMEDUC, RILLI, GRSA, PPWATER, PPWC, PPGARB and PPENERGY, average values for the variables PAREAPERM, PMCFSJ, PC714FE, PINWORK,

INTIND, IDHMLONG, ESPLIFE and MPRODCOCO and low values for variables TOTCOX, TOTBIRDS, TRACTORS, PPOOR, PCI, RFECUN, PPDOWN, P714A and PPIBAGRI.

It was observed that the municipalities of the High Hinterland region form a more homogeneous, cohesive and well defined region, while the Central Wasteland Territory, as well as the Low San Francisco, have some variation degree which could imply the existence of subdivisions. This suggests that policy makers should pay attention to internal territorial differences and particularities and decide according to the best solution of each subdivision. Other neural map territories have shown good degree of cohesion by the Component Planes analysis.

### **Automatic Clustering by TerraSOM**

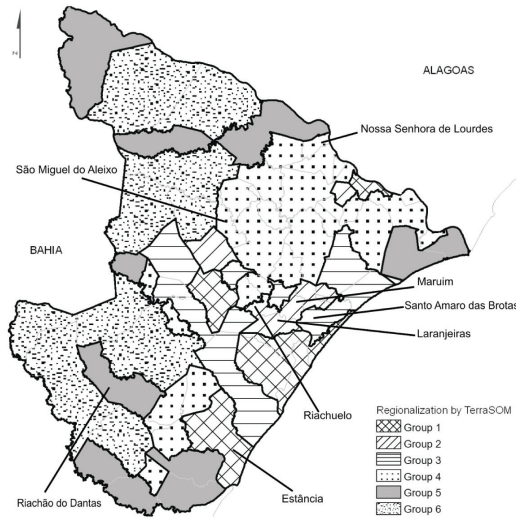
From the distribution of the municipalities on the neural map and the variables analysis by the Component Planes it was possible to propose a new arrangement of the Territories of Identity. Although the neural map regionalization by visual interpretation was not conclusive, it was possible to proceed with a third clustering by the Costa-Netto algorithm to improve the amount of information about the aggregation of Sergipe's municipalities and validate the previous results.

Figure 6 shows the regionalization of the state of Sergipe by a one-dimensional SOM, with 10 neurons and linearly initiated. Several network sizes were tested, but basically the same groups were found, varying not more than in terms of aggregation or subdivision of each one. The automatic clustering algorithm found 6 groups, as shown in Figure 6. Important to note that there was no contiguity restriction on the clustering algorithm, so non-neighboring municipalities clusters were allowed.

The automatic regionalization confirms the previous results. The municipality of Riachão do Dantas belongs to the group 5, the same group of municipalities in the South Territory (Tomar do



Figure 6. Automatic regionalization by the TerraSOM plugin without neighborhood constraints



Geru, Indiaroba, Santa Luzia do Itanhý and Cris-tinópolis). The municipality of Estancia belongs to the group 1, the same group of the municipa-lities of the Great Aracaju Territory. Also note the distinction between these municipalities in the groups 2 and 3, which belong to the Grand Aracaju Territory. Nevertheless, according to the SOM network, they belong to the East Territory (Laranjeiras, Riachuelo, Maruim and Santo Amaro das Brotas).

It was also confirmed by the automatic regiona-lization that the municipalities of Nossa Senhora de Lourdes and São Miguel do Aleixo were closer in the feature space with the Middle Hinterland Territory. It was observed that the group 4 con-gregated municipalities from three Territories of Identity (Lower San Francisco, Central wasteland and Middle Hinterland), which demonstrates the difficulty of separating these municipalities.

### The Final Regional Partition

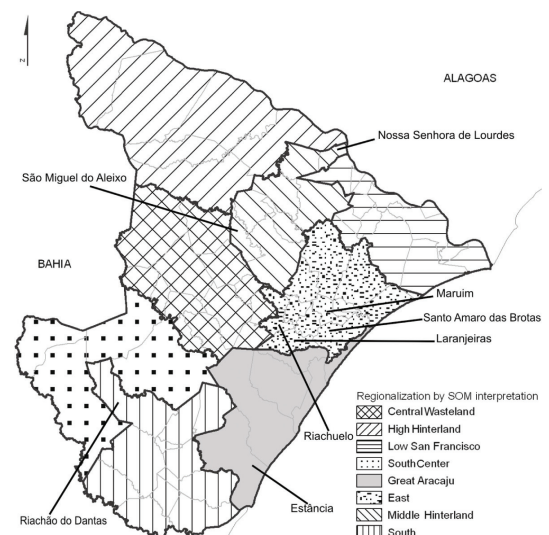
After the visual and automatic clustering and inter-pretation of the neural map a new regionalization of the state of Sergipe was proposed (Figure 7).

It was taken into account the contiguity among municipalities during the process of construction of the new partition. The consistency of the results of the three forms of the SOM analysis improved the consistency of the proposal.

Figure 7 shows the final configuration of the state of Sergipe partition based on the forty-four variables. The municipality of Riachão do Dantas migrated from the South Central Territory to the South Territory, while Estância migrated from the Southern Territory to the Great Aracaju Territory. The municipalities of Laranjeiras, Maruim, Santo Amaro das Brotas and Riachuelo migrated from the Great Aracaju Territory to the Eastern Territory. The municipality of São Miguel do Aleixo migrated from the Central Wasteland Territory to the Middle Hinterland Territory. And finally, the city of Nossa Senhora de Lourdes migrated from the High Hinterland Territory to the Middle Hinterland Territory.

This final regionalization proposal is not intended to replace the official one, but rather demonstrates the ability to undertake the territo-rialization from an exploratory spatial analysis by a neural network. Teixeira et al. (2007) evidenced

Figure 7. Proposed regionalization by SOM algorithm



that several key criteria to describe the association to one or another territory is also intimately related to group and personal abstract values, e.g., Nossa Senhora de Lourdes (LOU) and São Miguel do Aleixo (SMA) are quantitatively within the reach of the Middle Hinterland, but the participative process resulted in a different association between these municipalities and the Territories of Identity.

Estância (EST), Riachão do Dantas (RDT) and Itaporanga D' Ajuda (IAJ) are vast municipalities, and the existence of concentrated activities could lead to misclassification. However, considering only the forty-four variables mentioned in this study, it was concluded that Estância is more related to the Great Aracaju Territory, once this municipality has a great industrial GDP and a low agricultural production. The geographical space of the RDT municipality goes from the Wasteland to the Hinterland, despite of the indication of a deeper relation to the South Territory, especially to the municipalities of the extreme south, as it can be observed in Figure 5. Maruim (MRI), Laranjeiras (LRJ), Santo Amaro das Brotas (SAB) and Riachuelo (RIA) present features nearer the East Territory than the Great Aracaju. The sugar cane activity, the governmental transferences (mainly the royalties), and the way of social organization may approximate this group to municipalities of the East Territory.

## **FUTURE RESEARCH DIRECTIONS**

The SOM acts as an extractor of features without reducing the size of the feature vector, i.e., without significant loss of statistical information on geospatial dataset. This allows the use of statistical methods associated with the neural network, especially in the post-processing of the codevectors  $w$ . Hierarchical clustering methods or algorithms based on vector quantization can be applied to the clustering process of the codevectors. Depending on the relationship between

the number of input vectors and the size of the feature vector, the application of the Principal Component Analysis (PCA) could be necessary for a dimensionality reduction.

There is also the possibility to add new components in the input dataset so that new factors could be considered in the process of regionalization. It is possible, for example, to include an explicit component of spatial location of the municipalities into the input vector or into the learning process of the SOM algorithm for a spatial constraint. Numerical and categorical data fusion can be performed for a better representation of different dimensions, both environmental and economical. Additional temporal information can be added for a trend analysis using the SOM's projection algorithms.

As an interdisciplinary study a lot of work is necessary to join distinct concepts and actions from many disciplines, such as economy, sociology, geography and agronomy towards rural territorial typology for a sustainable development. However, this integrative and systemic approach is still a challenger for researchers and policy makers.

## **CONCLUSION**

The study of the sustainable development of rural areas depends on the growth of integrated and multidisciplinary studies. The regional approach has been one of the methods to design projects of development and regional research in order to consider social, economic, geophysical and cultural dimensions. Population growth, and hence the demand for quality public services and projects for regional infrastructure, require from public managers greater flexibility in decision-making, so that they observe the requirements of fairness, efficiency and effectiveness. Territorial typology certainly provides important feedback for the development of a collective solution to issues relating to rural development.

The large volume of data already available and their differences in origin and format imposes research challenges, requiring new methods of data analysis, especially those related to data mining. The possibility of visualization of multivariate data combined with the ability to use automatic clustering algorithms turned the neural network SOM into an important tool in the process of territorial typology.

Considering forty-four heterogeneous variables the Territory of Identity presented by the dataset distribution on artificial neural map intra and inter spatial dependence and low level of inter-territorial shifting of municipalities (only 12%), however, the analysis by Component Planes have shown some subdivision intra-territories showing that policy makers must be careful about simple and homogeneous solutions for each Territory. The automatic regionalization by the SOM algorithm integrated with the Costa-Netto partition strategy associated with the CDBw validity cluster index confirmed that there are significant differences intra-territories and that there are no-neighbor municipalities with similar characteristics (for example, Riachão do Dantas, Canindé do São Francisco and Brejo Grande). The automatic clustering has also shown that some Centralities (Itabaliana, Estância and Propriá) are closer to the Great Aracaju Territory and that some municipalities of the Great Aracaju Territory could be associated to another group or to the East Territory.

Based on the Territories of Identity and on the results of the SOM algorithm was constructed a regionalization for rural studies purposes. It can help the general understanding about the Sergipes's rural actual scenario and aid the process of regional planning. This is particularly important for regions where agriculture is a major employer of labor, family labor and, therefore, important source of income for the farming populations.

## REFERENCES

- Abramovay, R. (2006). Para una teoría de los estudios territoriales. In Manzanal, M., Neiman, G., & Lattuada, M. (Eds.), *Desarrollo rural – Organizaciones, instituciones y territorios* (pp. 51–70). Buenos Aires, Argentina: Ediciones Ciccus.
- Aksoy, E. (2006, October) *Clustering with GIS: an attempt to classify turkish district data*. Paper presented at the XXIII FIG Congress. Munich, Germany.
- Baçon, F., Lobo, V., & Painho, M. (2005). The self-organizing map, the Geo-SOM, and relevant variants for geosciences. *Computers & Geosciences*, 31, 155–163. doi:10.1016/j.cageo.2004.06.013
- Bailey, T. C., & Gatrell, A. C. (1995). *Interactive Spatial Data Analysis*. Longman.
- Bandeira, P. S. (2006). Institucionalização de regiões no Brasil. *Revista da Sociedade Brasileira para o Progresso da Ciência*, 58(1), 34-37.
- Barreto, G. A., & Araújo, A. F. R. (2001). Time in self-organizing maps: An overview of models. *International Journal of Computer Research*, 10(2), 139–179.
- Claval, P. (2008). Espaces et territoire: les bifurcations de la science régionale. *Géographie, Économie, Société*, 10, 157–184.
- Costa, J. A. F. (1999) *Classificação automática e análise de dados por redes neurais auto-organizáveis*. Unpublished doctoral dissertation, University of Campinas, Campinas, São Paulo.
- Costa, J. A. F., & Andrade Netto, M. L. (2003). *Segmentação do SOM Baseada em Particionamento de Grafos*. Paper presented at the VI Brazilian Congress on Neural Networks, São Paulo, (pp. 451-456).

- Flexer, A. (2001). On the use of self-organizing maps for clustering and visualization. *Intelligent Data Analysis*, 5, 373–384.
- Flores, M. (2004). Desenvolvimento territorial rural: uma proposta de estudo para apoio à formulação de políticas públicas. In Lages, et al (Ed.), *Territórios em movimento: cultura e identidade como estratégia de inserção competitiva*. Brasília, DF: SEBRAE.
- Genovez, P. (2002). *Território e Desigualdades: análise espacial intra-urbana no estudo da dinâmica de exclusão/inclusão social no espaço urbano em São José dos Campos-SP*. Unpublished master thesis, National Institute for Space Research, São José dos Campos, São Paulo, Brazil.
- Haese, K., & Goodhill, G. J. (2001). Auto-SOM: Recursive Parameter Estimation for Guidance of Self-Organizing Feature Maps. *Neural Computation*, 13, 595–619. doi:10.1162/089976601300014475
- Halkidi, M., & Vazirgiannis, M. (2002). *Clustering validity assessment using multi representatives*. Paper presented at the SETN Conference, Thessaloniki, Greece.
- Kaski, S., & Kohonen, T. (1996). Exploratory data analysis by the self-organizing map: structures of welfare and poverty in the world. In Refenes, A.-P. N., Abu-Mostafa, Y., Moody, J., & Weigend, A. (Ed.), *Third International Conference on Neural Networks in the Capital Markets* (pp. 498-507). Singapore: World Scientific.
- Kohonen, T. (2001). *Self-organizing maps*. Berlin: Springer.
- Koua, E. L. (2003). *Using self-organizing maps for information visualization and knowledge discovery in complex geospatial datasets*. Paper presented at the XXI International Cartographic Conference (ICC), (pp. 1694-1702).
- Ministério do Desenvolvimento Agrário – MDA. (2005). *Referências para a gestão social de territórios rurais*. Brasília, DF: MDA. Série Documentos nº 3.
- Openshaw, S., & Turton, I. (1996). A parallel Kohonen algorithm for the classification of large spatial datasets. *Computers & Geosciences*, 22(9), 1019–1026. doi:10.1016/S0098-3004(96)00040-4
- Sabourin, E., & Teixeira, O. A. (2002). *Planejamento e desenvolvimento dos territórios rurais*. Brasília: Embrapa Informação Tecnológica.
- Silva, M. A. S., Monteiro, A. M. V., & de Medeiros, J. S. (2004a). *Visualization of geospatial data by component planes and u-matrix*. Paper presented at the VI Brazilian Symposium on GeoInformatics, Campos do Jordão, Brazil.
- Silva, M. A. S., Monteiro, A. M. V., & de Medeiros, J. S. (2004b). *Semi-Automatic Geospatial Data Clustering by Self-Organizing Maps*. Paper presented at the VI Brazilian Symposium on Neural Networks, 2004, São Luis, MA.
- Spielman, S. E., & Thill, J.-C. (2008). Social area analysis, data mining, and GIS. *Computers, Environment and Urban Systems*, 32, 110–122. doi:10.1016/j.compenvurbsys.2007.11.004
- Teixeira, O. A., Melo, R. L. de & França, V. L. A. (2007). *Estado de Sergipe: uma proposta de territorialização para o planejamento*. SEPLAN Report.
- Varsta, M. (2002). *Self-organizing maps in sequence processing*. Unpublished doctoral dissertation, Helsinki University of Technology. Espoo, Finland.
- Vesanto, J. (1997). *Data Mining Techniques Based on the Self-Organizing Map*. Unpublished master thesis, Helsinki University of Technology.

Vesanto, J. (1999). SOM based data visualization methods. *Intelligent Data Analysis*, 3, 111–126. doi:10.1016/S1088-467X(99)00013-X

Vesanto, J., & Alhoniemi, E. (2000). Clustering of the Self-Organizing Map. *IEEE Transactions on Neural Networks*, 11(3), 586–600. doi:10.1109/72.846731

Winter, K., & Hewitson, B. (1994). Selforganizing maps - applications to census data. In Hewitson, B., & Crane, R. (Eds.), *Neural nets: applications in geography*. Amsterdam: Kluwer.

Wu, S., & Chow, T. W. (2004). Clustering of the self-organizing map using a clustering validity index based on inter-cluster and intra-cluster density. *Pattern Recognition*, 37, 175–188. doi:10.1016/S0031-3203(03)00237-1

## **ADDITIONAL READING**

Babu, G. P. (1997). Self-organizing neural networks for spatial data. *Pattern Recognition Letters*, 18, 133–142. doi:10.1016/S0167-8655(97)00003-2

Bishop, C. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press.

Cai, Y. (1995). Artificial neural-network method for soil-erosion forecasting. *Bodenkultur*, 46(1), 19–24.

Cereghino, R., Giraudel, J., & Compin, A. (2001). Spatial analysis of stream invertebrates distribution in the Adour-Garonne drainage basin (France), using Kohonen self organizing maps. *Ecological Modelling*, 146, 1–3, 167–180. doi:10.1016/S0304-3800(01)00304-0

Couclelis, H. (1998). Geocomputation in context. In Longley, P., Brooks, S., McDonnell, R., & Macmillan, B. (Eds.), *Geocomputation: a primer*. John Wiley and Sons.

Demartines, P., & Blayo, F. (1992). Kohonen self-organizing maps: is the normalization necessary? *Complex Systems*, 6, 105–123.

Erwin, E., Obermayer, K. K., & Schuler, K. (1992). Self-Organizing maps: Stationary states, metastability and convergence rate. *Biological Cybernetics*, 67, 35–45. doi:10.1007/BF00201800

Fausett, L. (1994). *Fundamentals Neural Networks: Architectures, algorithms, and applications*. Prentice Hall.

Fischer, M., & Getis, A. (1996). *Recent developments in spatial analysis*. Springer.

Foody, G. (1999). Applications of the self-organising feature map neural network in community data analysis. *Ecological Modelling*, 120, 97–107. doi:10.1016/S0304-3800(99)00094-0

Gahegan, M. (1999). What is Geocomputation? *Transactions in GIS*, 3, 203–206. doi:10.1111/1467-9671.00017

Gahegan, M., Takatsuka, M., Wheeler, M., & Hardisty, H. (2002). Introducing GeoVISTA Studio: an integrated suite of visualization and computational methods for exploration and knowledge construction in geography. *Computers, Environment and Urban Systems*, 26, 267–292. doi:10.1016/S0198-9715(01)00046-1

Haese, K. (1998). Self-Organizing Feature Maps with Self-Adjusting Learning Parameters. *IEEE Transactions on Neural Networks*, 9, 1270–1278. doi:10.1109/72.728376

Hewitson, B., & Crane, R. (1994). *Neural nets: applications in geography*. Kluwer.

Ji, C. Y. (2000). Land-use classification of remotely sensed data using self-organizing feature map neural networks. *Photogrammetric Engineering and Remote Sensing*, 66(12), 1451–1460.

Kaski, S., Kangas, J., & Kohonen, T. (1998). Bibliography of Self-Organizing Map (SOM) Papers: 1981–1997. *Neural Computing Surveys*, 1, 102–350.

Koga, D. (2003). *Medidas de Cidades: entre territórios de vida e territórios vividos*. Cortez.

Kohonen, T. (1990). The Self-organizing maps. *Proceedings of the IEEE*, 78(9), 1464–1480. doi:10.1109/5.58325

Kropp, J. (1998). A neural network approach to the analysis of city systems. *Applied Geography (Sevenoaks, England)*, 18(1), 83–96. doi:10.1016/S0143-6228(97)00048-9

Lo, Z., Yu, Y., & Bavarian, B. (1993). Analysis of the convergence properties of topology preserving neural networks. *IEEE Transactions on Neural Networks*, 4, 207–220. doi:10.1109/72.207609

Muñoz, A., & Muruzábal, J. (1998). Self-organizing maps for outlier detection. *Neurocomputing*, 18, 33–60. doi:10.1016/S0925-2312(97)00068-4

Oja, M., Kaski, S., & Kohonen, T. (2003). Bibliography of Self-Organizing Map (SOM) Papers: 1998–2001 Addendum. *Neural Computing Surveys*, 3, 1–156.

Openshaw, S., & Abrahart, S. (2000). *Geocomputation*. Taylor & Francis.

Openshaw, S., & Openshaw, C. (1997). *Artificial intelligence in geography*. John Wiley.

Purvis, M., Zhou, Q., Cranefield, S., Ward, R., Raykov, R., & Jessberger, D. (2001). Spatial information modelling and analysis in a distributed environment. *Ecological Modelling & Software*, 16, 439–445. doi:10.1016/S1364-8152(01)00014-7

## KEY TERMS AND DEFINITIONS

**Artificial Neural Network (ANN):** ANN is a computational model that simulates functional aspects of biological neural networks. It consists of an interconnected set of artificial neurons and processes information using a parallel approach to computation. ANN is usually used to find patterns in datasets.

**Best Match Unit (BMU):** BMU is the neuron  $j$  closer to some input  $x_k$  than other neurons. It can also be called winner neuron.

**Codevectors:** Codevectors are the set of  $w_j$ ,  $j=1, 2, \dots, m$ ,  $m$  is the number of neurons.

**Learning Algorithm:** Algorithm that allows computers to evolve behaviors based on empirical data, such as from sensor data or databases.

**Neural Topology:** Neural topology refers to the layout of connected neurons.

**Quantization:** Quantization is the process of mapping a large set of possible discrete values (or a continuous range of values) by a relatively small and finite set of discrete values which is also on that original interval.

**Voronoi Region:** Voronoi region is a kind of a metric space determined by Euclidean distances to a specified discrete set of objects (codevectors  $w_j$ ). Each Voronoi region is formed by all input vectors  $x_k$  closer to each reference object ( $w_j$ ) than any other codevector.

## ENDNOTE

- <sup>1</sup> Municipalities that concentrate services and opportunities for the rest of the same Territory of Identity.