

Abordagem Conexionista para Análise Espacial Exploratória de Dados Socioeconômicos de Territórios Rurais

Marcos Aurélio Santos da Silva¹
Edmar Ramos de Siqueira²
Olívio Alberto Teixeira³

Resumo: O uso de técnicas de análise multivariada para avaliação espacial exploratória de dados agregados por áreas rurais tem se mostrado útil na caracterização e tipologia territorial. Ganha destaque nesta tarefa a abordagem conexionista, em especial a rede neural não supervisionada do tipo Mapa Auto-Organizável (MAO), ao agregar funções de análise de agrupamentos e de visualização de dados. Este trabalho estudou a capacidade da rede MAO no processo de avaliação exploratória de dados oriundos de múltiplas fontes para fins de determinação de agregados locais no Território Rural Centro-Sul de Sergipe (TRCSS). Os resultados evidenciaram que, apesar de homogêneo, o TRCSS apresenta subgrupos territoriais com diferenças significativas para as variáveis estudadas.

Palavras-chave: mapa auto-organizável, visualização, agrupamento, território.

Abstract: *The use of multivariate analysis techniques for exploratory spatial study of rural geospatial data has being useful in territorial characterization and typology. Connectionist approach is a very efficient computational method for these tasks, mainly the Self-Organizing Map (SOM). Through the use of SOM clustering and data visualization can be conducted simultaneously. This paper assesses the capacity of SOM algorithm in the task of exploratory multisource data analysis for local aggregation verification in the Center-South Rural Territory of Sergipe State, Brazil. Results shown that, despite the homogeneity of the region, there are significant differences among sub-groups of municipalities that compose the territory.*

Key-words: *self-organizing maps, visualization, clustering, territory.*

Classificação JEL: C630, R120.

¹ E-mail: aurelio@cpatc.embrapa.br

² E-mail: edmar@cpatc.embrapa.br

³ E-mail: teixeira@ufs.br

1. Introdução

Desde 2003, importantes aspectos das políticas públicas de desenvolvimento rural brasileiro foram definidos a partir de recortes territoriais. Assim, a definição de território rural adotada pelo Ministério do Desenvolvimento Agrário (MDA) é clara quanto à multidimensionalidade dos fatores constituintes e formadores desse espaço (MDA, 2005). Embora a coesão sociocultural seja um dos pilares da identidade regional, componentes biofísicos também auxiliam na ‘delimitação’ do território, ao determinar a fonte de recursos naturais e econômicos que afetarão o padrão de ocupação espacial. Vale ressaltar a forte tendência de concentração populacional em ambientes favoráveis (com oferta hídrica, qualidade dos solos, acesso a mercados etc.).

Para o MDA, outra etapa fundamental na construção de territórios consiste na definição dos conceitos de empoderamento e governança. O primeiro diz respeito à tomada de consciência por meio do acesso a informações sobre território, cidadania, direitos e deveres, bem como da maior capacidade de comunicação em fóruns colegiados representativos e democráticos. Uma boa governança denota coesão social, através da diversidade de representação nos ambientes decisórios, com amplo controle social a partir da tomada de decisão colegiada bem articulada, com base em demandas, fatos e dados reais. Observa-se, portanto, que a oferta de informação estruturada e organizada sobre o território constitui-se fator determinante no estudo e entendimento do mesmo, além de auxiliar os processos de empoderamento e governança (SABOURIN e TEIXEIRA, 2002).

Este trabalho avaliou a presença ou ausência de agregados locais no Território Rural Centro-Sul de Sergipe⁴ (TRCSS), assim como tendências de aproximação e distanciamento de municípios no espaço de atributos para os períodos de referência 1996 e 2006. A abordagem conexcionista foi adotada para a análise de agregados territoriais, devido a suas características básicas de reconhecimento de padrões complexos em dados reais (HAYKIN, 2001). A rede neural não supervisionada Mapa Auto-Organizável (MAO) aplicada neste trabalho também combina a capacidade de análise de agrupamentos com mecanismos de visualização de dados (KOHONEN, 2001).

⁴ Ao final de 2008, a SDT/MDA alterou a composição do TRCSS de forma que os municípios de Lagarto e Riachão do Dantas migraram para o Território Sertão Ocidental de Sergipe, São Cristóvão também foi excluído do TRCSS. Apesar da alteração, o estudo foi mantido para o TRCSS, atualmente denominado de Território Sul de Sergipe.

2. Mapas Auto-Organizáveis

O Mapa Auto-Organizável (MAO) consiste num algoritmo de quantização vetorial e ordenação de dados, motivado por estudos do comportamento do córtex cerebral (KOHONEN, 2001). A rede MAO organiza os sinais de entrada (dados) num reticulado de neurônios uni, bi ou tridimensional, associados a vetores de código $w_j = [w_{j1}, w_{j2}, \dots, w_{jd}]^T$, que representarão o conjunto de dados de entrada da rede. Esse ordenamento possibilita a visualização dos dados e a consequente partição dos mesmos em grupos homogêneos em termos dos atributos considerados.

O reticulado hexagonal, $M \times N$, representado pelos m ($m = M \cdot N$) neurônios, constitui a arquitetura básica da rede MAO. O vetor de variáveis aleatórias $x_k = [\xi_{k1}, \xi_{k2}, \dots, \xi_{kd}]^T$ corresponde aos sinais de entrada da rede. Um conjunto de n instâncias de x_k deve ser apresentado à rede MAO para que, após o seu processamento, essas instâncias sejam representadas pelos vetores de código w de forma ordenada no reticulado.

Tem-se, a seguir, o algoritmo da rede MAO, considerando aprendizagem em lote:

- 1: iniciação linear dos vetores de código w ;
- 2: definição dos parâmetros de processamento da rede MAO;
 - 2.1: número de iterações de aprendizagem T , $t(0) = 1$;
 - 2.2: dimensões M e N da rede MAO;
 - 2.3: raio inicial, $\delta(0)$, do processo de atualização dos vetores de código.
- 3: para cada tempo t discreto e natural até T faça;
 - 3.1: para todo x_k encontre o neurônio c que seja o mais similar a x_k segundo a distância euclidiana (d_{ij}), $c = \operatorname{argmin}_j \{ \|w_k - w_j\| \}$, $j = 1, 2, \dots, m$, sendo m correspondente ao número de neurônios da rede.
 - 3.2: atualize os vetores de código do neurônio c e de seus vizinhos segundo o raio δ conforme as equações:

$$h_{ij}(t) = \exp(-d_{ij}^2 / 2\delta(t)^2) \quad (1)$$

$$s_i(t) = \sum_j^{m_i} x_j \quad (2)$$

$$w_i(t) = \frac{\sum_j^m h_{ij}(t) s_j(t)}{\sum_j^m n_{vj} h_{ji}(t)} \quad (3)$$

em que s_i representa o somatório dos padrões relativos à região de Voronoi V_i e n_{Vi} corresponde ao número de amostras do conjunto de Voronoi do neurônio i .

O termo $h(t)$ é a função que determina a vizinhança entre o neurônio vencedor c e seus vizinhos.

$$\delta(t) = \delta(t - 1) \exp(-t/\tau_i) \quad (4)$$

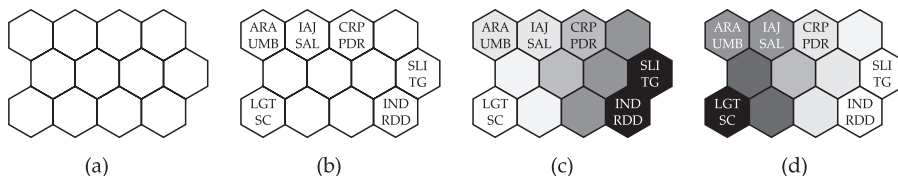
Após o processamento do algoritmo MAO, tem-se que a rede apresenta as seguintes propriedades:

- i) cada instância do vetor de entrada x_k está associada a um único neurônio;
- ii) cada vetor de código w_j codifica ou representa o conjunto de dados associados ao neurônio j ;
- iii) a proximidade física entre os neurônios e seus dados associados denota também proximidade no espaço de atributos. Todavia, pode haver neurônios vizinhos que não sejam similares entre si.

São essas propriedades que permitem a detecção de agrupamentos nos dados. Os métodos tradicionais de análise como o k-médias e a abordagem hierárquica sobre os vetores de código têm sido bastante aplicados (VESANTO e ALHONIEMI, 2000). No entanto, as análises das projeções dos vetores de código, para o caso das redes com reticulado bidimensional, se mostram mais efetivas na detecção de agrupamentos com estruturas complexas (ULTSCH, 1993; KASKI e KOHONEN, 1996; VESANTO, 1997; VESANTO, 1999).

A forma de projeção adotada neste trabalho baseou-se nos Planos de Componentes (PC's) (KOHONEN, 2001), gerados, um para cada variável, a partir dos vetores de código w . Gera-se, então, um padrão coroplético para a rede MAO, no qual, para cada variável i , se colore um neurônio j de acordo com o valor w_{ji} do vetor de código. A Figura 1 mostra as etapas de geração de dois planos componentes relativos ao processamento de uma rede MAO 2D, hexagonal e de dimensões 3x4 (Figura 1a). Processaram-se 12 vetores de entrada de dimensão igual a dois com posterior rotulação (Figura 1b). Foram gerados os PC's para as variáveis índice de pobreza e IDHM-Renda (Figura 1c e d).

Figura 1. Rede MAO 2D, hexagonal e de dimensões 3x4 antes do processamento (a). Rede neural rotulada após o processamento (b). PC's para as variáveis índice de pobreza e IDHM-Renda (c) e (d).



Fonte: Elaborada pelos autores.

A interpretação dos planos de componentes (PC) é bastante intuitiva. Cada PC corresponde a uma distribuição dos valores dos vetores representativos da

rede MAO. O intervalo de valores assumidos por w , para cada componente, será convertido em escalas de cinza segundo o seguinte padrão: o valor mais baixo é representado pela cor branca, e os demais valores, em ordem crescente, assumem tonalidades de cinza até que o valor mais alto seja representado pela cor preta. Para cada rede MAO, têm-se d PC's, sendo que d corresponde à dimensão do vetor x_k . A partir dos PC's, é possível verificar a presença de dados atípicos (MUÑOZ e MURUZÁBAL, 1998), avaliar correlação entre variáveis (VESANTO e AHOLA, 1999), além de estudar a dependência espacial dos componentes em análise (BAÇÃO et al., 2008; SPIELMAN e THILL, 2008).

3. Análise de dados geoespaciais a partir dos Mapas Auto-Organizáveis (MAO)

Embora a rede MAO esteja se consolidando como uma ferramenta eficiente de análise exploratória de dados, poucos são os trabalhos relacionados à dependência espacial (OPENSHAW, BLAKE e WYMER, 1994; SPIELMAN e THILL, 2008). Considerando que não há uma teoria para validar a primeira lei de Tobler (1970), que afirma que localidades próximas são mais parecidas que as que estão distantes umas das outras, muitas pesquisas empíricas sugerem formas de abordagem prática para esse problema.

Kaski e Kohonen (1996) elaboraram um dos primeiros estudos sobre o uso dos planos de componentes para avaliação de agregados territoriais, a partir de informações estatísticas de um determinado conjunto de países. Apesar de os dados estarem associados a localizações geográficas, não foi efetuada análise alguma de dependência espacial. Já o trabalho de Winter e Hewitson (1994) procurou, mesmo que de forma restrita, usar os PC's como mecanismo de avaliação da distribuição espacial, com base em dados demográficos. Openshaw e Openshaw (1997) enfatizaram a capacidade da rede MAO na análise espacial, destacando, sobretudo, as possibilidades de uso de variados métodos de visualização dos dados.

Trabalhos mais recentes apontam para o crescimento de técnicas e métodos vinculados à rede MAO para a análise da dependência espacial. Bação et al. (2005) incluem explicitamente no algoritmo de processamento da rede as coordenadas geográficas dos dados multivariados oriundos do censo demográfico. Esse procedimento força a existência de relação de proximidade dos espaços físicos e de atributos, mas impede a verificação exploratória da dependência espacial.

A análise temporal, neste trabalho, foi realizada a partir do uso de duas redes MAO para interpretação dos dois conjuntos de dados de referência: de 1996 e 2006. Skupin e Hagelman (2005) avaliaram a fusão dos dados temporais num único conjunto, para serem interpretados através dos PC's de redes com $m \gg n$ e uso de interpoladores sobre os PC's, para posterior interpretação visual das trajetórias das unidades de estudo no mapa neural. Embora seja efetiva, essa

estratégia deve ser aplicada após a análise dos métodos básicos de visualização de dados com a rede MAO.

Spielman e Thill (2008) elaboraram estudo bastante conclusivo sobre generalização descritiva de dados geodemográficos, com alto nível de acoplamento entre o algoritmo MAO e o Sistema de Informação Geográfica (SIG), usado para geração dos mapas síntese. Porém, os PC's não foram usados para a análise dos dados. Os autores inverteram o processo de associação, partindo do mapa físico para o mapa neural. Essa técnica, apesar de eficiente, depende do desenvolvimento de ferramentas de ligação entre a rede MAO e o SIG.

Neste trabalho, adotou-se a rede MAO bidimensional 20x15, com reticulado hexagonal, função de vizinhança gaussiana, iniciação linear dos vetores de código, algoritmo de aprendizagem em lote unifásico com 1.000 iterações e raio inicial de atualização dos vetores de código igual a 15. O pacote SOMToolbox (VESANTO et al., 1999) foi usado para processamento dos dados.

4. O Território Rural Centro-Sul de Sergipe (TRCSS)

A composição inicial do Território Rural Centro-Sul de Sergipe (TRCSS) continha 15 municípios: Tomar do Geru (TG), Umbaúba (UMB), São Cristóvão (SC), Itabaianinha (ITN), Salgado (SAL), Estância (EST), Riachão do Dantas (RDD), Cristinápolis (CRP), Itaporanga D'Ajuda (IAJ), Indiaroba (IND), Santa Luzia do Itanhy (SLI), Pedrinhas (PDR), Lagarto (LGT), Arauá (ARA) e Boquim (BUI).

O Território ocupa uma área de 5.441,80 km², com as principais atividades econômicas da região concentradas na agricultura. Há forte participação das lavouras permanentes (citros e coco), das pastagens (destaca-se o rebanho bovino), além de expressiva presença da agricultura familiar.

Para a análise exploratória de agregados regionais no TRCSS e a avaliação de tendência entre os períodos de referência 1996 e 2006, considerou-se como população de estudo todos os municípios do estado (74), exceto Aracaju, e 34 variáveis oriundas dos censos agropecuários do IBGE (Instituto Brasileiro de Geografia e Estatística) de 1996 (IBGE, 2009a) e 2006 (IBGE, 2008), as médias aritméticas da Produção Agrícola Municipal de 1991-1996 e de 2001-2006 (IBGE, 2009b) e informações de 1991 e 2000 relativas ao Atlas de Desenvolvimento Humano no Brasil, do Programa das Nações Unidas para o Desenvolvimento (PNUD, 2009), divididas em nove fatores, conforme a Tabela 1.

O município de Aracaju foi retirado da população por ter características bastante distintas dos demais. Apresenta valores extremos para a maioria das variáveis, o que aproximaria as outras cidades no espaço de atributos, impossibilitando a visualização dos dados no mapa neural de forma clara.

A escolha das variáveis baseou-se na disponibilidade de informações sobre os períodos próximos a 1996 e a 2006, referenciais dos censos agropecuários. Como há diferenças entre as variáveis adotadas entre os dois censos e, considerando

que o IBGE liberou somente dados preliminares do censo 2006, as componentes dos censos restringiram-se aos fatores uso da terra, tratores, efetivo de animais e produção animal. Os dados sobre produção vegetal vieram da PAM/IBGE (Pesquisa Agrícola Municipal do IBGE), sendo que, para o ano referência 1996, adotou-se a média da produção agrícola entre 1991 e 1996 e, analogamente, para o ano referência 2006, a média da produção agrícola entre 2001 e 2006.

Tabela 1. Fatores e variáveis usadas na avaliação do território rural.

| Fator | N | Variável |
|--------------------|----|---|
| Uso da terra | 1 | % do município ocupado com lavouras |
| | 2 | % do município ocupado com pastagens |
| | 3 | % do município ocupado com matas |
| Tratores | 4 | Total de tratores para cada 100 estabelecimentos |
| Efetivo de animais | 5 | Total de bovinos por estabelecimento |
| | 6 | Total de caprinos por estabelecimento |
| | 7 | Total de ovinos por estabelecimento |
| | 8 | Total de suínos por estabelecimento |
| | 9 | Total de aves por estabelecimento |
| Produção animal | 10 | Quantidade (mil litros) de leite de vaca por estabelecimento |
| | 11 | Quantidade de ovos (mil dúzias) de galinha para cada 100 estabelecimentos |
| Educação | 12 | IDHM-Educação 1991, 2000 |
| | 13 | Taxa de alfabetização 1991, 2000 |
| | 14 | Taxa bruta de frequência à escola 1991, 2000 |
| | 15 | Taxa bruta de frequência ao fundamental 1991, 2000 |
| | 16 | Taxa bruta de frequência ao ensino médio 1991, 2000 |
| | 17 | Taxa bruta de frequência ao superior 1991, 2000 |
| Longevidade | 18 | IDHM Longevidade 1991, 2000 |
| | 19 | Esperança de vida ao nascer 1991, 2000 |
| Renda | 20 | IDHM Renda 1991, 2000 |
| | 21 | Renda per Capita 1991, 2000 |
| Pobreza | 22 | Percentual de indigentes 1991, 2000 |
| | 23 | Percentual de pobres 1991, 2000 |
| | 24 | Percentual de crianças indigentes 1991, 2000 |
| | 25 | Percentual de crianças pobres 1991, 2000 |
| | 26 | Intensidade da pobreza 1991, 2000 |
| | 27 | Intensidade da indigência 1991, 2000 |
| Produção vegetal | 28 | Média da produção de milho 1991-1996, 2001-2006 |
| | 29 | Média da produção de mandioca 1991-1996, 2001-2006 |
| | 30 | Média da produção de feijão 1991-1996, 2001-2006 |
| | 31 | Média da produção de cana 1991-1996, 2001-2006 |
| | 32 | Média da produção de laranja 1991-1996, 2001-2006 |
| | 33 | Média da produção de banana 1991-1996, 2001-2006 |
| | 34 | Média da produção de coco 1991-1996, 2001-2006 |

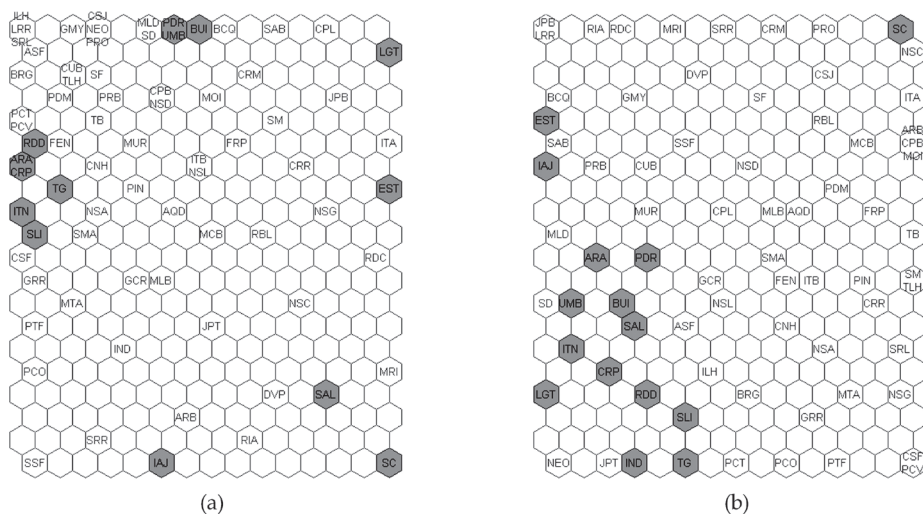
Fonte: Elaborada pelos autores.

Considerando o objetivo de incluir no processo de análise variáveis não relacionadas com fatores eminentemente agrícolas e, levando em conta a restrição da necessidade de comparação temporal, adotaram-se variáveis referentes aos fatores educação, longevidade, renda e pobreza do Atlas de Desenvolvimento Humano no Brasil/PNUD. Para o ano referência 1996, foram utilizados os valores relativos ao ano de 1991 e, da mesma forma, para o ano referência 2006, os valores de 2000.

5. Resultados e discussão

Após o processamento da rede MAO, para cada ano referência (1996 e 2006), rotularam-se os neurônios com as siglas municipais, de forma que a proximidade no espaço de atributos entre dois municípios pudesse ser observada através da proximidade física no reticulado neural. Assim, em 1996, constatou-se dispersão no espaço de atributos entre os municípios do TRCSS, com a consequente presença de cidades isoladas como LGT, EST, SAL, SC e IAJ. Já em 2006, observou-se uma aproximação entre os municípios do TRCSS, exceto SC, que permaneceu isolado das demais cidades do território (Figura 2a e 2b).

Figura 2. Rede neural após associação dos neurônios com os municípios com maior proximidade entre o vetor de características e o vetor de código do neurônio.



(a) Configuração final para os dados do ano referência 1996; (b) Configuração final para os dados do ano referência 2006; Municípios do TRCSS em destaque.

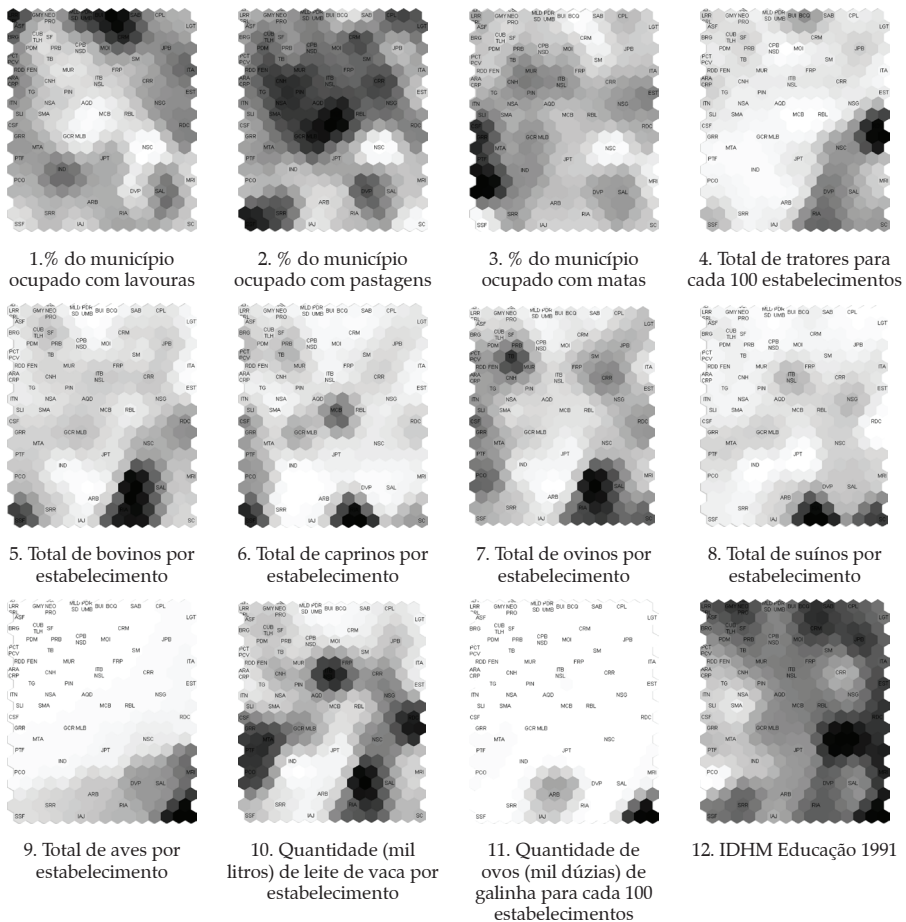
Fonte: Elaborada pelos autores.

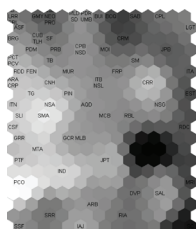
A tendência à homogeneização no espaço de atributos dos municípios do TRCSS é um fator extremamente relevante quando se tem como propósito a

elaboração de políticas públicas regionais de desenvolvimento. No entanto, a visualização da distribuição dos municípios no mapa neural não informa em que direção estará indo esse processo de equiparação. Para fazer a avaliação qualitativa da distribuição das cidades, conforme os valores das variáveis (através da visualização dos vetores de código), gerou-se os PC's.

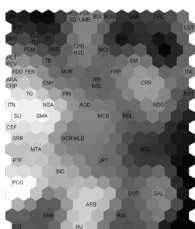
São 34 planos de componentes para cada ano referência, conforme as Figuras 3 e 4. Embora todos os municípios estejam representados nos PC's, far-se-á referência somente àqueles do território em análise. O grande número de PC's impede que todos sejam apresentados com os rótulos perfeitamente visíveis. Logo, recomenda-se consulta à Figura 2 para associação do neurônio com o município.

Figura 3. Planos de componentes das 34 variáveis relacionadas ao ano referência de 1996.

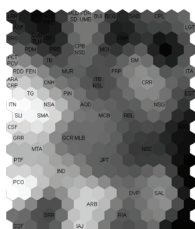




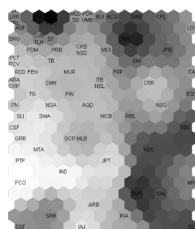
13. Taxa de alfabetização 1991



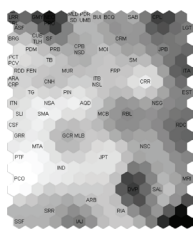
14. Taxa bruta de frequência à escola 1991



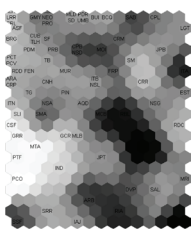
15. Taxa bruta de frequência ao fundamental 1991



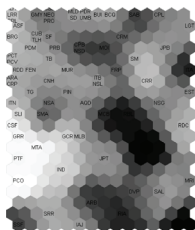
16. Taxa bruta de frequência ao ensino médio 1991



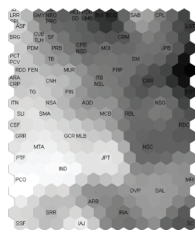
17. Taxa bruta de frequência ao superior 1991



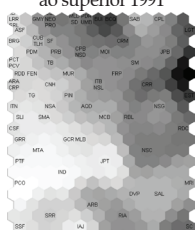
18. IDHM Longevidade 1991



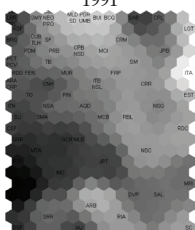
19. Esperança de vida ao nascer 1991



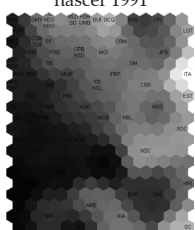
20. IDHM Renda 1991



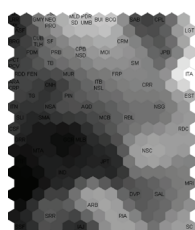
21. Renda per Capita 1991



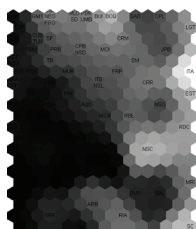
22. Percentual de indigentes 1991



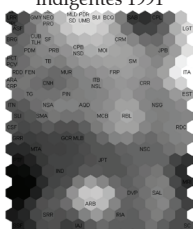
23. Percentual de pobres 1991



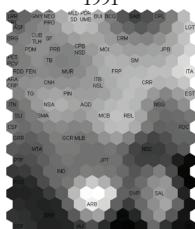
24. Percentual de crianças indigentes 1991



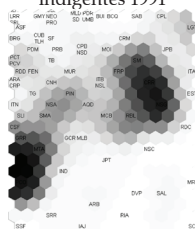
25. Percentual de crianças pobres 1991



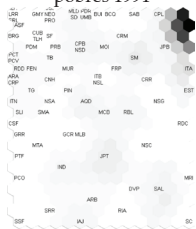
26. Intensidade da pobreza 1991



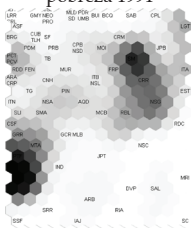
27. Intensidade da indigência 1991



28. Média da produção de milho 1991-1996



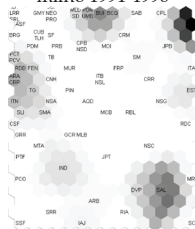
29. Média da produção de mandioca 1991-1996



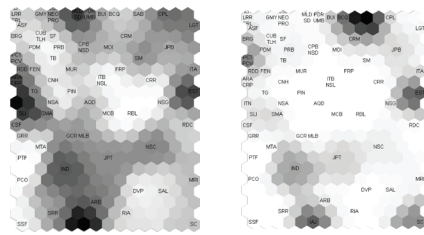
30. Média da produção de feijão 1991-1996



31. Média da produção de cana 1991-1996



32. Média da produção de laranja 1991-1996



33. Média da produção de banana 1991-1996

34. Média da produção de coco 1991-1996

Fonte: Elaborada pelos autores.

Para os dados de referência de 1996 (planos de componentes da Figura 3), observa-se predomínio de municípios com grande percentual do seu território coberto por pastagens. Destaque para o reduzido número de cidades com grande percentual da área coberta por lavouras, incluindo PDR, UMB e BUI.

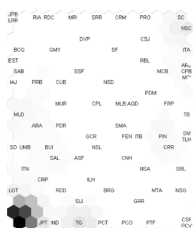
Quanto ao fator efetivo de animais, verifica-se correlação entre as variáveis efetivo de bovinos e ovinos e entre as variáveis efetivo de caprinos e suínos. São Cristóvão (SC) apresenta valores de efetivo de suínos e aves bem acima da média dos demais municípios. Para o fator produção animal, destaca-se novamente SC, com produção de ovos bastante superior à média do estado.

Os PC's relativos ao fator educação mostram claramente correlação entre IDHM-Educação e taxa de alfabetização, entre taxa bruta de frequência à escola e taxa bruta de frequência ao fundamental, e entre as variáveis taxa bruta de frequência ao ensino médio e taxa bruta de frequência ao superior. Os municípios de LGT, EST, SAL e SC apresentaram os melhores índices para o fator educação, sendo que SC registrou, ainda, os melhores índices relacionados ao fator longevidade.

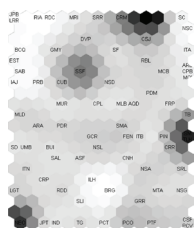
As duas variáveis referentes à renda tiveram correlação. Os municípios de EST, LGT, BUI, PDR e UMB apresentaram os melhores índices relativos à renda. O fator pobreza mostrou que há correlação entre as variáveis percentual de indigentes e intensidade da pobreza e entre as variáveis percentual de pobres, percentual de crianças indigentes e percentual de crianças pobres. Os melhores índices para esse fator foram registrados pelos municípios de EST, LGT, PDR, UMB e BUI. Para o fator produção vegetal, as variáveis obtiveram alto grau de homogeneidade, com destaque para a produção de mandioca em LGT. Para esse fator, o TRCSS mostrou vocação para a produção de banana (todos os municípios do TRCSS) e coco (EST e IAJ).

Quanto às variáveis do ano referência 2006 (Figura 4), observa-se dados atípicos para o fator uso da terra, que não apresenta relação alguma com o TRCSS. O fator efetivo de animais mostrou correlação entre as variáveis efetivo de aves e suínos e entre o efetivo de caprinos e ovinos. O município de SC registrou grande concentração de suínos e aves. Para o fator produção animal, destacou-se novamente o SC, com a maior produção de ovos do estado.

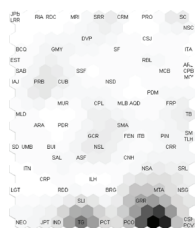
Figura 4. Planos de componentes das 34 variáveis relacionadas ao ano referência de 2006.



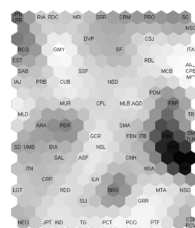
1. % do município ocupado com lavouras



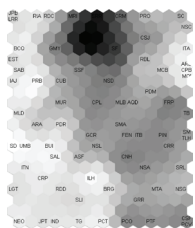
2. % do município ocupado com pastagens



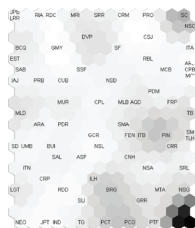
3. % do município ocupado com matas



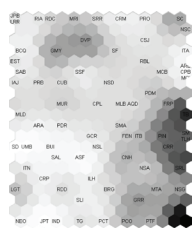
4. Total de tratores para cada 100 estabelecimentos



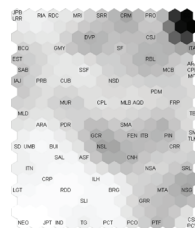
5. Total de bovinos por estabelecimento



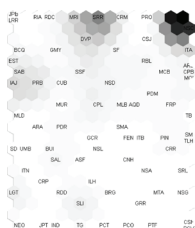
6. Total de caprinos por estabelecimento



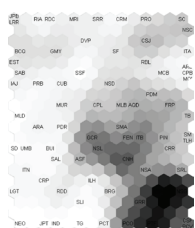
7. Total de ovinos por estabelecimento



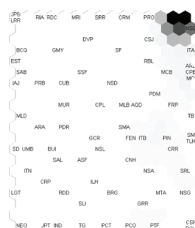
8. Total de suínos por estabelecimento



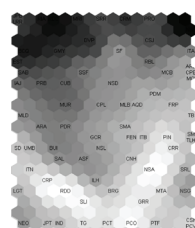
9. Total de aves por estabelecimento



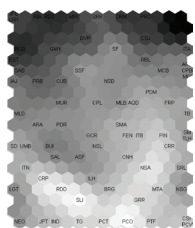
10. Quantidade (mil litros) de leite de vaca por estabelecimento



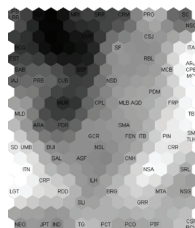
11. Quantidade de ovos (mil dúzias) de galinha para cada 100 estabelecimentos



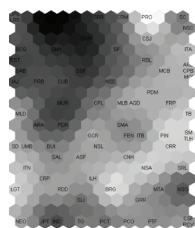
12. IDHM Educação 2000



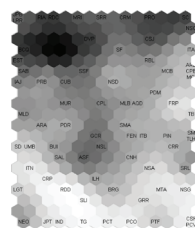
13. Taxa de alfabetização 2000



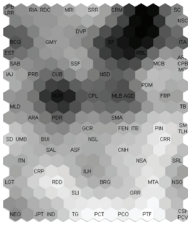
14. Taxa bruta de frequência à escola 2000



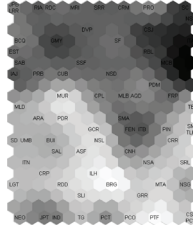
15. Taxa bruta de frequência ao fundamental 2000



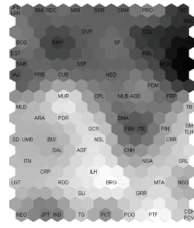
16. Taxa bruta de frequência ao ensino médio 2000



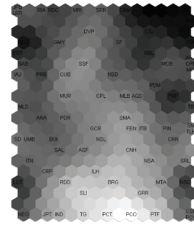
17. Taxa bruta de frequência ao superior 2000



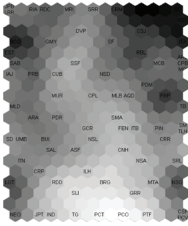
18. IDHM Longevidade 2000



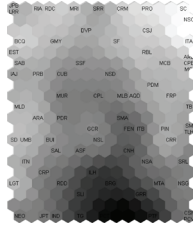
19. Esperança de vida ao nascer 2000



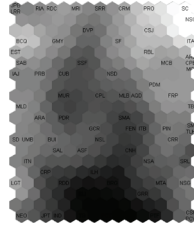
20. IDHM Renda 2000



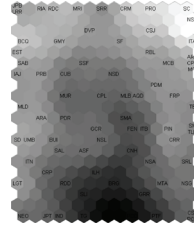
21. Renda per Capita 2000



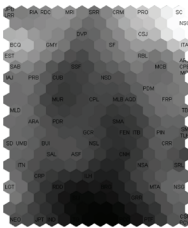
22. Percentual de indigentes 2000



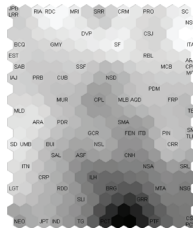
23. Percentual de pobres 2000



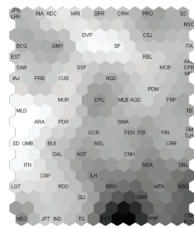
24. Percentual de crianças indigentes 2000



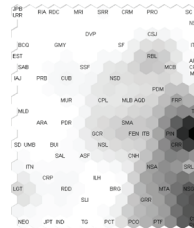
25. Percentual de crianças pobres 2000



26. Intensidade da pobreza 2000



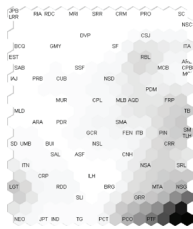
27. Intensidade da indigência 2000



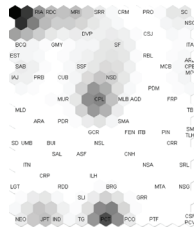
28. Média da produção de milho 2001-2006



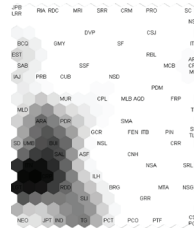
29. Média da produção de mandioca 2001-2006



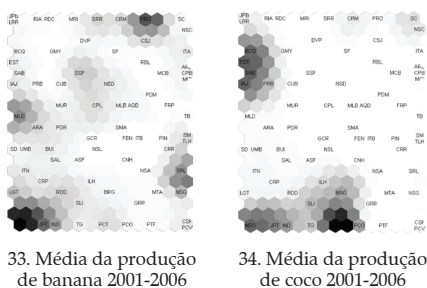
30. Média da produção de feijão 2001-2006



31. Média da produção de cana 2001-2006



32. Média da produção de laranja 2001-2006



Fonte: Elaborada pelos autores.

Para o fator educação, houve correlação entre as variáveis IDHM-Educação e taxa de alfabetização e entre taxa bruta de frequência à escola e taxa bruta de frequência ao fundamental. Os municípios de EST, IAJ e SC registraram os melhores índices para esse fator. As duas variáveis relacionadas ao fator longevidade apresentaram correlação. Todos os municípios do TRCSS obtiveram baixa longevidade, exceto SC.

As variáveis do fator renda tiveram correlação, sendo que a maioria dos municípios do TRCSS apresentou índices de médio para alto, exceto IND, RDD, SLI e TG. Sobre o fator pobreza, houve correlação entre as variáveis percentual de indigentes, percentual de pobres, percentual de crianças indigentes e percentual de crianças pobres, e entre as variáveis intensidade da pobreza e intensidade da indigência. O município de SC registrou os melhores índices para esse fator; ARA, UMB, EST, IND e LGT apresentaram resultados intermediários; e as demais cidades do TRCSS obtiveram dados bem inferiores.

Para o fator produção vegetal, as variáveis apontaram alto grau de homogeneidade, com destaque para a produção de mandioca em LGT. Para esse fator, o TRCSS mostrou vocação para a produção de laranja (todos os municípios do TRCSS) e coco (EST e IAJ).

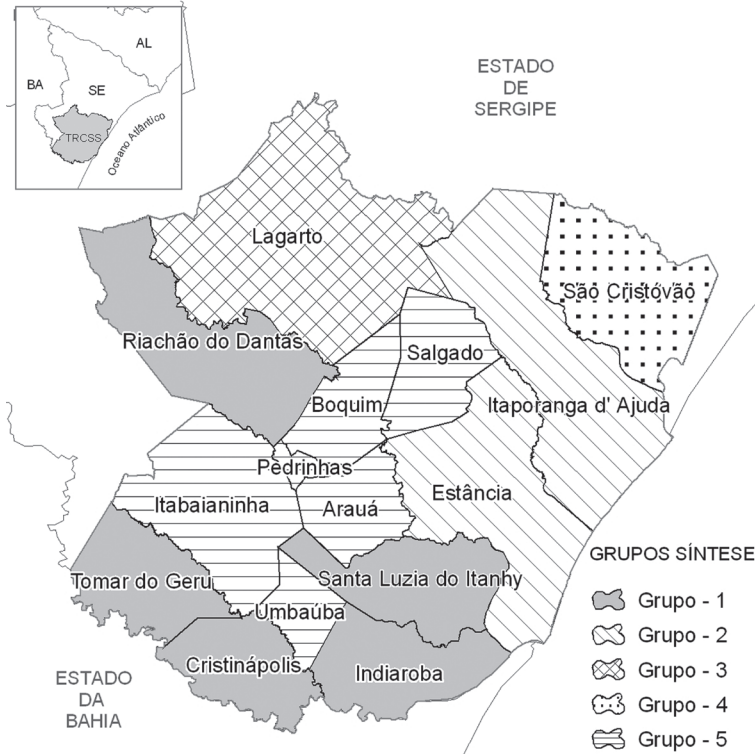
A análise dos planos de componentes apontou que a dispersão no espaço de atributos em 1996 está relacionada à maior variação das variáveis dos fatores educação, longevidade, renda e pobreza. Para os fatores vinculados à atividade agropecuária (uso da terra, efetivo de animais, produção animal e produção vegetal), observou-se menor variação em 1996, tendendo para maior homogeneidade em 2006.

São Cristóvão destaca-se dos demais municípios do TRCSS em praticamente todos os fatores avaliados. Os seus indicadores não agropecuários estão muito próximos dos daqueles municípios da Grande Aracaju, como Nossa Senhora do Socorro e Aracaju.

Quando se comparam os planos de componentes relativos à renda e pobreza, verifica-se que, para os dados de referência 2006, houve avanços no número de municípios com indicadores positivos para ambos os fatores. No

TRCSS, observa-se crescimento da renda e diminuição da pobreza em CRP, ITN, SC, ARA e SAL. Em BUI, UMB e PDR, nota-se pequena contração da renda e leve aumento da pobreza. Os demais municípios permaneceram relativamente no mesmo patamar em relação ao estado.

Figura 5. Subagregados territoriais no TRCSS.



Fonte: Elaborada pelos autores.

Segundo a distribuição no espaço de atributos (rede MAO) dos municípios, a partir dos dados de referência 2006, tem-se a seguinte partição (Figura 5): IND, TG, SLI, RDD e CRP formariam o grupo 1, caracterizado por elevados índices de pobreza, baixos níveis educacionais, de renda e de longevidade. Os municípios de EST e IAJ comporiam um segundo grupo, marcado por baixos índices de pobreza, elevados índices educacionais, de longevidade e de renda. Os municípios de LGT e SC formariam dois grupos unitários com características bem distintas, enquanto SC destaca-se pela presença de valores extremos e bons níveis para os indicadores sociais, LGT apresenta-se mais próximo das demais cidades do TRCSS, embora se diferencie levemente em alguns aspectos. Por fim,

ARA, PDR, BUI, SAL, UMB e ITN formariam um agrupamento de municípios caracterizados por apresentarem-se entre os dois extremos (grupos 1 e 2).

6. Conclusões

O Mapa Auto-Organizável (MAO) de Kohonen mostrou-se eficaz na tarefa de análise espacial exploratória de dados multivariados. A partir da sua aplicação, foi possível identificar possibilidades de correlação entre variáveis, detectar dados atípicos, analisar tendência na aproximação/distanciamento de municípios, bem como avaliar agregados locais com indicação de dependência espacial.

O TRCSS configura-se como um território homogêneo, embora se tenha constatado que, para as variáveis analisadas, o município de São Cristóvão mostrou-se bastante distante dos demais. Cinco subagregados (Figura 5) locais, com forte dependência espacial, foram identificados pela análise dos planos de componentes, o que mostra a necessidade de refinamento das políticas públicas direcionadas ao território.

Vale destacar que, com a publicação do IBGE das demais componentes do censo agropecuário de 2006, será possível realizar a agregação de mais variáveis ao processo analítico. Dessa maneira, a análise regional poderá ser incrementada a partir da inclusão de novos fatores em futuros estudos, dentre os quais se destacam: técnicas de cultivo, investimentos, perfil do pessoal ocupado no campo e assistência técnica.

Por fim, espera-se que o método de análise exploratória de dados multivariados geoespaciais exposto neste trabalho possa colaborar no processo de caracterização dos Territórios Rurais e da Cidadania já definidos pelo Ministério do Desenvolvimento Agrário (MDA), bem como auxiliar na construção de outros recortes regionais para efeito de planejamento e elaboração de políticas públicas.

7. Referências Bibliográficas

BAÇÃO, F., LOBO, V.; PAINHO, M. The self-organizing map, the Geo-SOM, and relevant variants for geosciences. *Computers & Geosciences*, v. 31, p. 155-163, 2005.

HAYKIN, S. *Redes neurais: princípios e práticas*. Porto Alegre: Bookman, 2001. 900 p.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). *Censo agropecuário 1996*. IBGE- Rio de Janeiro: IBGE - Sistema IBGE de recuperação automática – SIDRA. Disponível: <http://www.ibge.gov.br> – acesso fev. 2009a.

_____. *Produção Agrícola Municipal*. Rio de Janeiro: IBGE - Sistema IBGE de recuperação automática – SIDRA. Disponível: <http://www.ibge.gov.br> – acesso fev. 2009b.

_____. *Censo Agropecuário - 2006 Resultados Preliminares*. Rio de Janeiro: IBGE, 2008.

KASKI, S.; KOHONEN, T. Exploratory data analysis by the Self-Organizing Map: structures of welfare and poverty in the world. In: *Neural Networks in the Capital Markets*, 3., 1996, London. *Proceedings...* London: World Scientific, 1996. p. 498–507.

KOHONEN, T. *Self-Organizing Maps*. Berlin: Springer, 2001. 501 p.

MINISTÉRIO DO DESENVOLVIMENTO AGRÁRIO (MDA). Referências para a gestão social de territórios rurais. Brasília, DF:MDA. Série Documentos nº 3, 2005.

MUÑOZ, A.; MURUZÁBAL, J. Self-Organizing Maps for outlier detection. *Neurocomputing*, v. 18, n. 1-3, p. 33–60, January 1998.

OPENSHAW, S.; BLAKE, M.; WYMER, C. *Using neurocomputing methods to classify britain's residential areas*. Leeds, 1994. (Working paper 94/17).

OPENSHAW, S.; OPENSHAW, C. *Artificial intelligence in geography*. Chichester: John Wiley and Sons, 1997. 348 p.

PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO (PNUD). Atlas do desenvolvimento humano no Brasil. Disponível em: <<http://www.pnud.org.br/atlas/>>. Acesso em: 15 jan. 2009.

SABOURIN, E.; TEIXEIRA, O.A. *Planejamento e desenvolvimento dos territórios rurais: conceitos, controvérsias, experiências*. Brasília: Embrapa, 2002

SKUPIN, A.; HAGELMAN, R. Visualizing demographic trajectories with self-organizing maps. *Geoinformatica*, v. 9, n. 2, 159-179, 2005.

SPIELMAN, S.E.; THILL, J.-C. Social area analysis, data mining and GIS. *Computers, Environment and Urban Systems*, v. 32, p. 110-122, 2008.

TOBLER, W.R. A computer movie simulating urban growth in the detroit region. *Economic Geography*, v.46, p.234-240, 1970.

ULTSCH, A. Knowledge extraction from self-organizing neural networks. In: Opitz, O. ed. *Information and Classification*. Berlin: Springer, 1993. p. 301–306.

VESANTO, J. *Data mining techniques based on the Self-Organizing Map*. 1997. 245 p. Dissertação (Mestrado em Computação) - Helsinki University of Technology, Helsinki. 1997.

_____. SOM based data visualization methods. *Intelligent Data Analysis*, v. 3, n. 2, p. 111–126, August 1999.

VESANTO, J.; AHOLA, J. Hunting for correlations in data using the Self-Organizing Map. In: International ICSC Congress on Computational Intelligence Methods and Applications (CIMA '99), 1999, Rochester, NY. *Proceedings...* Rochester, NY: ICSC Academic Press, 1999. p. 279–285.

VESANTO, J.; ALHONIEMI, E. Clustering of the Self-Organizing Map. *IEEE Transactions on Neural Networks*, v. 11, n. 3, p. 586–600, May 2000.

VESANTO, J.; HIMBERG, J.; ALHONIEMI, E.; PARHANKANGAS, J. Self-Organizing Map in matlab: the SOM toolbox. In: Matlab DSP Conference, 1999, Espoo, Finland. *Proceedings...* Espoo, Finland: Comsol Oy, 1999. p. 35–40.

WINTER, K.; HEWITSON, B.C. Self-organizing maps: applications to census data. In: HEWITSON, B.C.; CRANE, R.G. *Neural nets: applications in geography*, Dordrecht: Kluwer, 1994.