

Regionalização da agropecuária paranaense por meio de redes neurais artificiais

Marcos Adami¹

André Lima¹

Ramon Morais de Freitas¹

Antonio Miguel Vieira Monteiro¹

Mauricio Alves Moreira¹

¹ Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12245-970 - São José dos Campos - SP, Brasil
{adami, andre, ramon, mauricio}@dsr.inpe.br; miguel@dpi.inpe.br

Abstract. This work presents a characterization of the agriculture of Paraná State based on twenty-five variables extracted from social, economic and environmental data. In this analysis all variables were integrated using the Self Organizing Map network architecture and GIS software. The Kohonen maps provided by the Self Organizing Map analysis showed correlated variables and spatial clusters. The results indicated that rural income per area is negatively correlated to the Gini index, and positively correlated with total available soil water. The other environmental and socio-economic variables did not show relationships with the rural income per area. The Kohonen maps indicated that all variables related to the Human Development Index are highly correlated. The inclusion of the spatial location variables showed that rural income per area changes in an East-West direction, indicating spatial dependence. The cluster analysis showed seven regions that are related with the land use and land cover.

Palavras-chave: self organizing map, neural network, spatial analysis, mapas auto-organizáveis, redes neurais, análise espacial.

1. Introdução

O agronegócio corresponde aproximadamente a 30% do PIB brasileiro. No período de 1990 a 2002 enquanto o PIB total cresceu a uma média de 2,71% a.a., o PIB do agronegócio cresceu a 3,18% (Gasques et al., 2004). Estes valores mostram que a o agronegócio cresceu 8% a mais do que a economia total brasileira nos 12 anos, o que denota o desempenho do setor em comparação a economia do país. Não obstante, há de se conhecer as relações pertinentes a este setor econômico para que a sociedade possa intervir de forma a assegurar o desenvolvimento.

Dentro da cadeia produtiva do agronegócio, o produtor rural representa o elo frágil deste setor, pois, corre a maioria dos riscos envolvidos nesta atividade, e raras vezes são formadores de preço. Para manutenção do produtor rural, o conhecimento das relações entre as mais diversas variáveis que atuam neste domínio é importante. Neste sentido, a aplicação de políticas públicas na agricultura como a pesquisa e a extensão rural, assim como as iniciativas dos agentes do setor privado e das organizações não governamentais devem estar calcadas nas características de cada região e dos municípios que a integram (Fuentes Llanillo et al., 2006).

Para Campanhola e Graziano da Silva (2000) qualquer iniciativa de planejamento local deve se iniciar pela regionalização sócio-econômica e ambiental dos recursos disponíveis e pelo zoneamento territorial de modo que se tenha uma ocupação organizada tanto local como regional. Um diagnóstico analítico e descritivo das regiões, revela-se uma etapa fundamental para apontar conhecimentos ao planejamento e avaliação de ações que visem minimizar ou erradicar as situações de pobreza rural e as desigualdades regionais que ocorrem em diferentes concentrações fundiárias (Schneider e Waquil, 2001).

Desta forma, o estudo em questão tem por objetivo fazer a análise exploratória entre as variáveis: renda rural, fisiográficas, econômicas e sociais nos municípios do Paraná por meio

de redes neurais artificiais – Mapas Auto-Organizáveis (SOM) de Kohonen e obter a regionalização do Estado com base nestas variáveis.

2. Redes neurais artificiais – Mapas Auto-Organizáveis (SOM) de Kohonen

As redes neurais artificiais são modelos computacionais que simulam a capacidade humana de resolver problemas. Em geral, estes modelos são algoritmos computacionais que procuram representar, de maneira bastante elementar, o mecanismo de funcionamento cerebral. Suas vantagens são as características de adaptabilidade, generalização e tolerância a ruídos, dentre outras (Rogers e Kabrisky, 1991; Haykin, 2001).

Existem diversos algoritmos de redes neurais artificiais, e dentre eles a rede SOM. Este algoritmo foi idealizado a partir da analogia com a região do córtex cerebral humano. Descobriu-se que nesta parte do cérebro determinados locais são reservados para atividades específicas e que, para uma determinada ativação cerebral, o grau de ativação dos neurônios diminuía à medida que se aumentava a distância da região de ativação inicial (Kohonen, 2001).

A rede neural artificial SOM é composta por duas camadas, a camada I de entrada e a U de saída. Cada entrada na rede corresponde a um vetor com parâmetros que serão mapeados nos neurônios. Cada neurônio da camada de saída possui um vetor associado ao vetor de entrada, possibilitando mapear desta forma a origem dos dados. Os parâmetros essenciais deste algoritmo são: i) espaço de entrada contínuo, de padrões de ativação, gerados de acordo com uma distribuição de probabilidade; ii) topologia na forma da grade de neurônios que define um espaço de saída discreto; iii) função de vizinhança variável no tempo que é definida em torno do neurônio vencedor; iv) parâmetro da taxa de aprendizagem com um dado valor inicial e que diminui gradualmente com o tempo, mas nunca chega à zero (Rogers e Kabrisky, 1991; Haykin, 2001; Silva, 2004).

Desde o seu surgimento, em 1982, a rede SOM vem sendo aplicada numa ampla variedade de problemas de economia, engenharia, medicina, dentre outros. Destacam-se as potencialidades de visualização de dados multivariados, análise de agrupamentos, mineração de dados, descoberta de conhecimento e compressão de dados (Kohonen, 2001).

A análise exploratória de dados por meio da rede SOM, deve-se seguir os seguintes estágios: i) a escolha do conjunto de dados; ii) o pré-processamento dos dados; iii) a parametrização da rede; iv) escolha do tamanho do mapa neural e v) a interpretação dos resultados (Kaski e Kohonen, 1996). O primeiro estágio está envolvido com o estudo em questão. O segundo estágio envolve a normalização dos dados, para que a rede não atribua maior peso para uma determinada variável. A parametrização da rede SOM envolve a escolha da medida de vizinhança entre os neurônios, forma de vizinhança, raio inicial e forma e número de épocas de aprendizagem. A escolha do tamanho do mapa neural envolve a quantificação dos erros de quantização e de topologia. A interpretação dos resultados envolve a análise dos planos de componentes (Silva, 2004).

3. Material e Métodos

A área de estudo está localizada na região Sul do Brasil entre as latitudes sul 22° 29' e 26° 46', longitudes oeste 48° 00' e 54° 38'. O relevo do estado do Paraná, segundo Maack (1981) apresenta três características principais: (i) as altitudes mais elevadas ocorrem na serra do mar; (ii) a partir da serra do mar as altitudes decrescem no sentido leste para oeste e de noroeste para sudoeste; e (iii) a maior parte do território paranaense está contido entre as altitudes de 300 a 600 m. Os principais agrupamentos de solo presentes no estado, segundo Ministério da Agricultura (1981), são: Cambisol, Latossolo, Litólico, Podzólico, Terra Roxa e Sedimentar.

Os softwares SPRING (Câmara et al., 1996) e MATLAB (Mathworks, 2004) foram utilizados para realizar os processamentos dos dados. Estes dados são descritos na Tabela 1.

Tabela 1 – Descrição dos dados utilizados no trabalho.

DESCRIÇÃO	LEGENDA	FONTE	ESPECIFICAÇÕES
Valor bruto da Produção/hectare	vbp_ha	adaptado de Andretta (2006)	Valor total gerado pela agropecuária municipal dividido pela sua área
Coefficiente de Gini	G	adaptado de IBGE (2006)	Concentração de terras por município, dado por quantidade de estabelecimentos e pela área dos mesmos em hectares (ha).
Percentual do Solo do Município com alta capacidade de uso	p_a	Ministério da Agricultura (1981)	Aptidão agrícola para todos os tipos de cultivo, sem restrições.
Percentual do Solo do Município com média capacidade de uso	p_b	Ministério da agricultura (1981)	Aptidão agrícola para pecuária e fruticultura, a agricultura pode ser realizada com restrições.
Percentual do Solo do Município com baixa capacidade de uso	p_c	Ministério da Agricultura (1981)	Sem aptidão para agricultura.
Capacidade de Armazenamento de água Disponível no solo	CAD	Ministério da Agricultura (1981)	Quantidade de água em milímetros por metro cúbico de solo.
Unidades paisagísticas	planalto	Maack (1981);	Compartimentação geomorfológica a que pertence o município;
Declividade média do município	Dec_media	Rabus <i>et al.</i> (2003).	Representada em percentual
Desvio padrão da declividade do município	dec_dp	Rabus <i>et al.</i> (2003).	Representada em percentual
Amplitude da declividade do município	dec_amp	Rabus <i>et al.</i> (2003).	Representada em percentual
Razão de dependência rural do município	rur_razdep	Ipardes (2006)	Razão entre a soma da população rural com mais de 65 e menor de 15 anos pela população entre 15 e 65 anos. Os dados utilizados para o cálculo referem-se ao ano de 2000.
População rural com mais de 65 anos	rur_65a	Ipardes (2006)	Tendo por base o ano de 2000.
População rural com menos de 14 anos	rur0_14a	Ipardes (2006)	Tendo por base o ano de 2000.
População rural entre 15 e 64 anos	rur15_64a	Ipardes (2006)	Tendo por base o ano de 2000.
Índice de desenvolvimento humano – longevidade	idh_l	Ipardes (2006)	Reflete entre outras coisas, as condições de saúde da população, medida pela esperança de vida ao nascer. Refere-se ao ano de 2000.
Índice de desenvolvimento humano – educação	IDH_E	Ipardes (2006)	Integra a taxa de alfabetização de adultos e a taxa combinada de matrícula nos níveis de ensino: fundamental, médio e superior. Refere-se ao ano de 2000.
Índice de desenvolvimento humano – renda	Idh_r	Ipardes (2006)	Medida pelo poder de compra da população, baseado no PIB per capita ajustado ao custo de vida local para torná-lo comparável entre países e regiões, através da metodologia conhecida como paridade do poder de compra. Refere-se ao ano de 2000
Índice de desenvolvimento humano municipal	idh_m	Ipardes (2006)	Razão simples entre os índices IDH_L, IDH_E e IDH_R. Refere-se ao ano de 2000.
Renda per capita	r_capta	Ipardes (2006)	Razão entre a renda total do município pelo número de habitantes. Refere-se ao ano de 2000.
Longevidade	long_a	Ipardes (2006)	Esperança de vida ao nascer. Refere-se ao ano de 2000.
Taxa de crescimento da população rural	tx_cr_rur	Ipardes (2006)	Relação entre o ano de 2000 e o de 1991
Expectativa de vida	Esp_vida	Ipardes (2006)	Para o ano de 2000.
Taxa de alfabetização	tx_alf	Ipardes (2006)	Para o ano de 2000.
Latitude em coordenadas planas	latitude		Latitude do ponto médio do município
Longitude em coordenadas planas	longitude		Longitude do ponto médio do município

Os dados de renda por unidade de área do setor agropecuário para o ano de 2004 foram obtidos do valor bruto da produção agropecuária do Paraná (Andretta, 2006). O valor total estimado para cada município foi dividido pela área municipal, em hectares, para se obter o valor médio por hectare, por município.

Os dados de aptidão dos solos e CAD foram adaptados do mapa de aptidão do uso do solo para o estado do Paraná (Ministério da Agricultura, 1981) e os dados de declividade foram estimados a partir dos dados Shuttle Radar Topography Mission - SRTM (Rabus et al., 2003). Já as unidades paisagísticas do estado do Paraná foram oriundas dos estudos de Maack (1981).

Os dados de número de estabelecimentos agropecuários utilizados derivaram do censo agropecuário de 1995/96 (IBGE, 2006). Com base nestes valores, calculou-se o coeficiente de Gini (G), que é um indicador de desigualdade, para a distribuição dos estabelecimentos agropecuários para cada município do estado. Após o censo agropecuário de 1995/96 foram criados mais 28 municípios no estado do Paraná, passando de 371 para 399, teve-se a necessidade de estimar o G para os novos municípios. Esta estimativa foi feita com base na interpretação visual, através de imagens GEOCOVER (NASA, 2006), da estrutura fundiária dos “novos” municípios e dos municípios que circundavam este. Sendo o G estimado para o “novo” município proveniente do município vizinho com maior semelhança de estrutura fundiária. O G é calculado pela seguinte função (Hu, 1995):

$$G = \left| 1 - \sum_{k=0}^{k=n-1} (X_{k+1} - X_k)(Y_{k+1} - Y_k) \right| \quad (1)$$

em que X representa a proporção acumulada da variável estabelecimento agropecuário e Y representa a proporção acumulada da variável área. Os passos metodológicos para análise exploratória dos dados podem ser visualizados na Figura 1.

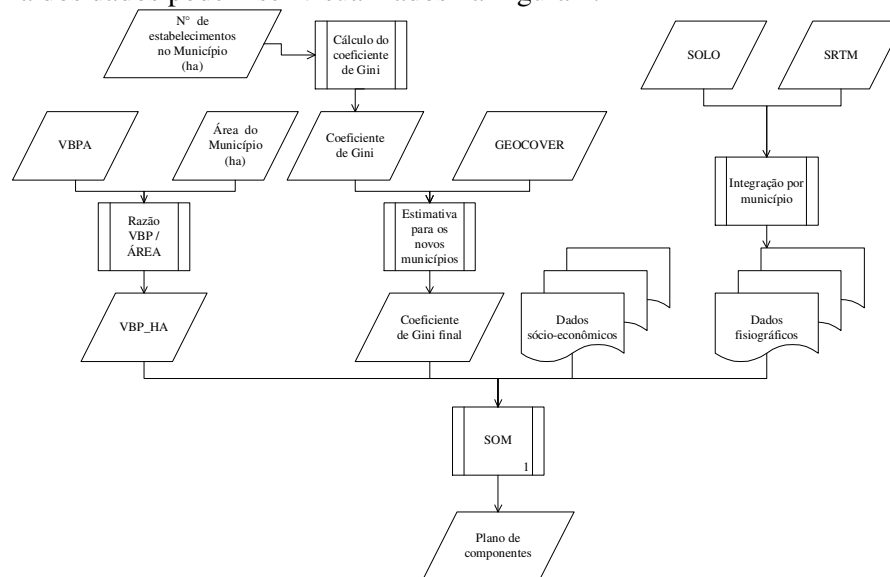


Figura 1 – Passos metodológicos até a obtenção do plano de componentes.

A análise exploratória dos dados foi realizada no MATLAB, por meio do toolbox SOM (Vessanto et al., 2000). Os dados de entrada na rede SOM foram normalizados [0-1] utilizando o histograma de cada variável. A definição da matriz de neurônios teve como ponto de partida o tamanho de vetores de entrada na rede. O tamanho mínimo de neurônios iniciais é dado pela quarta parte do número de vetores. O tamanho máximo é dado por quatro vezes o número de vetores. A definição da quantidade de neurônios final se dá quando o algoritmo minimiza os erros de quantização e topológico. Foi aplicada a SOM bidimensional, hexagonal, com função de vizinhança gaussiana e aprendizagem por lote. A seqüência de treinamento foi realizada em duas fases: a primeira fase, composta por 3 etapas, com um raio de vizinhança maior (5, 3, 1) e a segunda fase, composta por 10 etapas, com raio menor (0.5). Os planos de componentes foram utilizados para verificar o comportamento das variáveis e suas correlações com o VBPA. Também foi utilizado o algoritmo K-medias para agrupar os neurônios. Desta análise de cluster, obteve-se a regionalização do Estado do Paraná.

4. Resultados e Discussão

A Figura 2 apresenta os planos de componentes sem a inclusão das coordenadas planas dos municípios. O plano de componentes indica que as variáveis VBP_HA e GINI_AREA são inversamente correlacionadas. As setas destacam as regiões em que a resposta entre estes planos de componentes tiveram resposta inversa. Esta correlação inversa indica que no Paraná, os municípios onde a distribuição de terras é mais concentrada (GINI_AREA maior), a renda por hectare é menor. Este resultado corrobora com o trabalho de Biswanger e Elgin (1989) que apontam como motivos para que a renda dos municípios com melhor distribuição de terra sejam maiores. Pode-se verificar que a CAD é correlacionada com a renda (VBP_HA), ou seja, solos com melhor retenção de água estão em municípios com melhor renda por unidade de área.

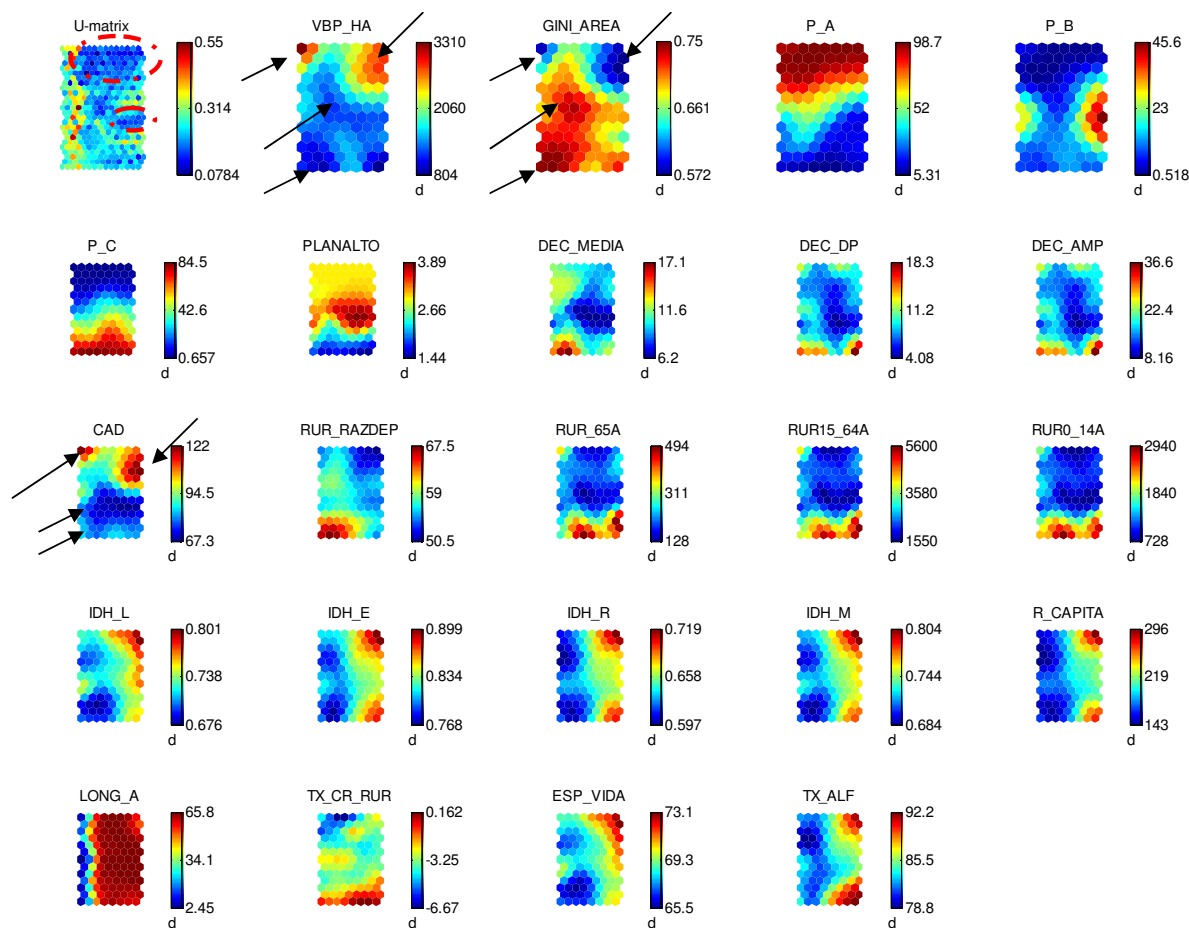


Figura 2 – Planos de componentes. As setas indicam as regiões de correlação entre as variáveis VBP_HA, GINI_AREA e CAD.

Os demais planos de componentes guardam pouca correlação com a variável estudada, o VBP_HA. Ao se estudar as correlações entre as demais variáveis verificar-se que o plano de componentes P_A é inversamente correlacionado com os planos P_B e P_C. O componente PLANALTO e DEC_MEDIA não apresentam correlação com os demais planos de componentes. Os componentes DEC_DP e DEC_AMP são correlacionados entre si. Os planos de componentes referentes à idade da população rural (RUR_65A, RUR0_14A e RUR15_64A) e são correlacionados e o plano de componentes IDH_L mostra relação com a ESP_VIDA, o IDH_E é função da TX_ALF e o IDH_R da variável R_CAPITA. A Figura 2 também apresenta a matriz U resultante. Nesta matriz há formação dos agrupamentos (detalhes em vermelho), região em que as distâncias entre os neurônios são baixas (cor azul).

A matriz U indica que são poucos os agrupamentos formados pelos neurônios, que relata grande variabilidade destas características nos municípios.

A Figura 3 apresenta os planos de componentes, com a inclusão das coordenadas planas, dos municípios. O plano de componentes mantém o mesmo padrão de correlações que foi observado anteriormente. Observa-se que o plano de componente LONGITUDE apresenta correlação com plano de componentes VBP_HA, o que denota uma dependência espacial da variável. Além disso, este plano de componentes indica que os municípios a oeste do estado geram maior quantidade de renda por unidade de área. Este fato não se repete quando se analisa a latitude, que não mostra que ocorrem diferenças entre a geração de renda, no sentido norte-sul. A questão da correlação inversa observada para o dado de LONGITUDE e a variável VBP deve-se a forma de exploração agropecuária. Em função da serra do Mar, ocorre baixa renda agropecuária na região Leste do Paraná, já na região Oeste, em função principalmente da suinocultura e da avicultura implantada na região de Francisco Beltrão, Cascavel e Toledo há um grande incremento do valor bruto da produção agropecuária (Andretta, 2006).

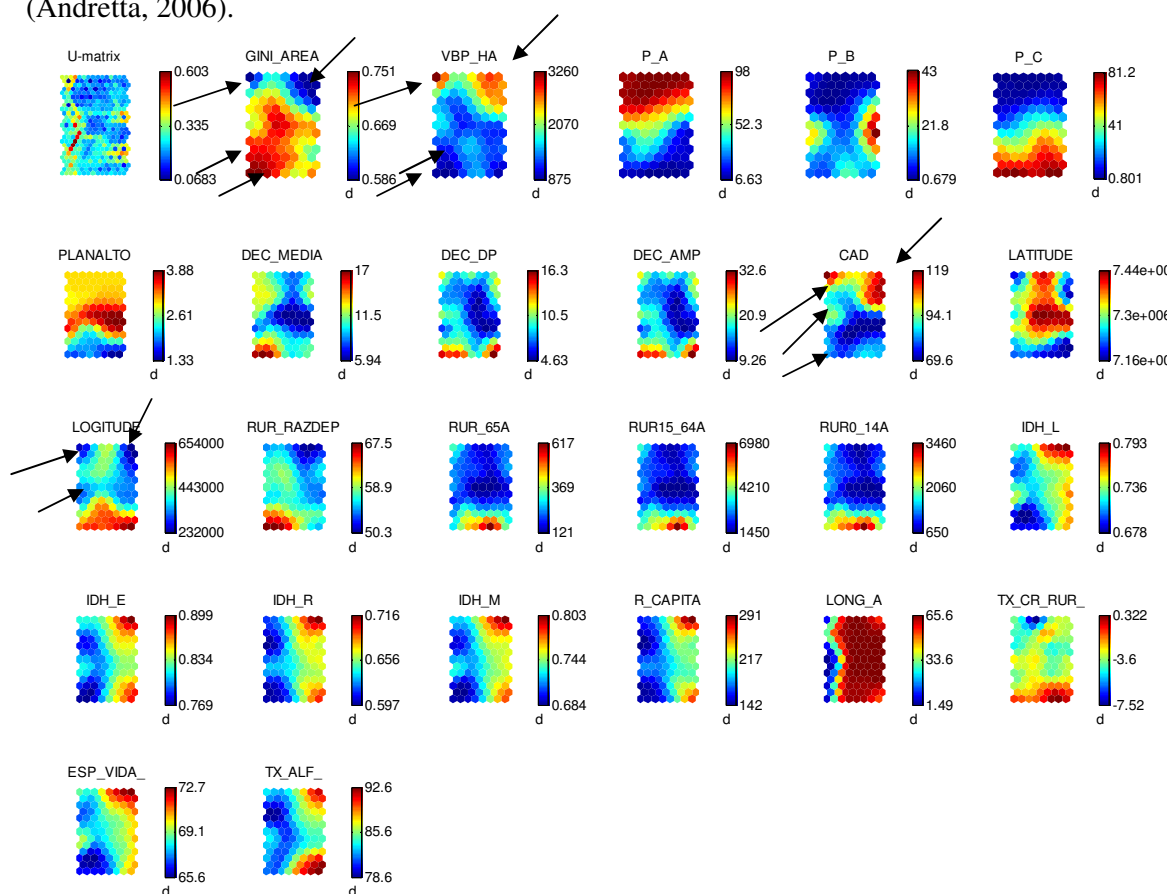


Figura 3 – Planos de componentes com inclusão da latitude e longitude.

Na Figura 4 são espacializadas as variáveis VBP_HA, GINI_AREA, CAD e a regionalização. Ao se analisar o comportamento espacial das variáveis VBP_HA (Figura 5a), GINI_AREA (Figura 5b) e CAD (Figura 5c) observa-se que as mesmas apresentam correlação, concordando com o plano de componentes. As setas indicam locais onde pode ser evidenciada esta correlação. A análise da correlação entre os componentes VBP_HA e GINI_AREA pode servir de auxílio a modelos de desenvolvimento econômico e na busca de soluções a questão fundiária. A aplicação do algoritmo K-medias nos neurônios resultou em 7

regiões, apresentadas na Figura 4d. As regiões 1 e 2 apresentam relações com a área destinada a agricultura, enquanto que a região 6 está associada a exploração da pecuária extensiva de corte. As relações obtidas entre as variáveis VBP_HA, GINI_AREA, CAD se mantêm na formação das regiões.

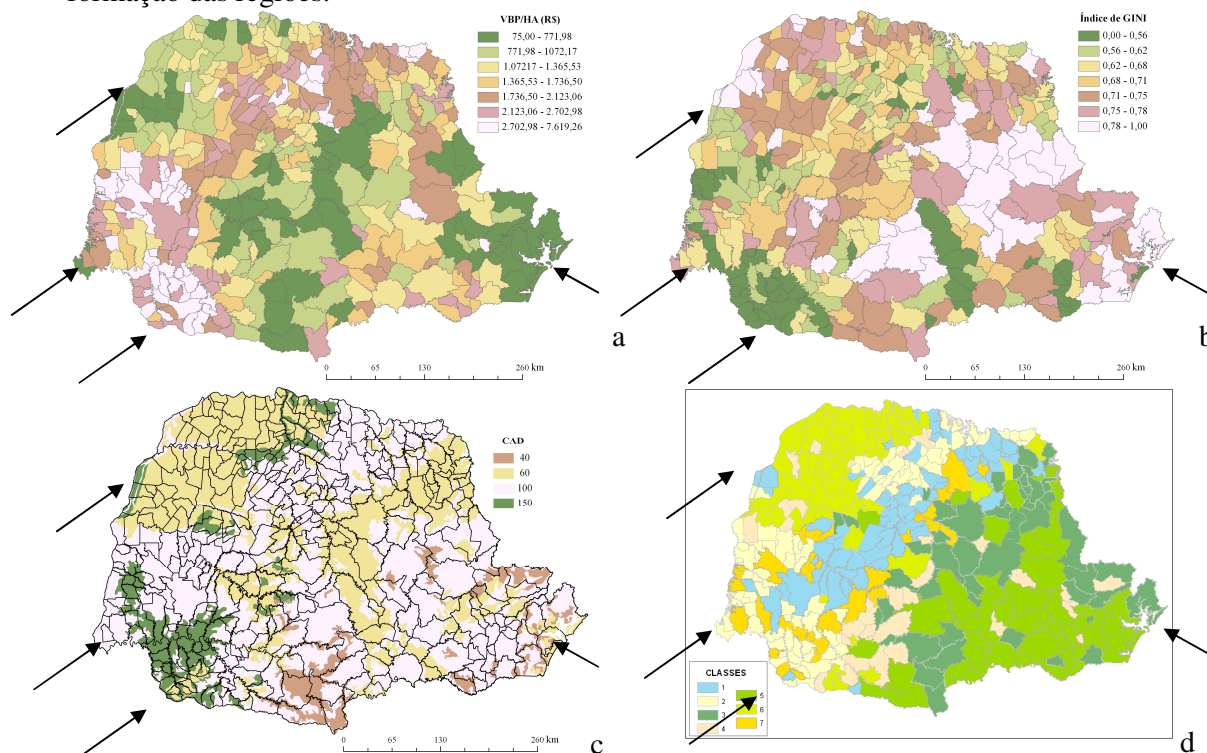


Figura 4 – Espacialização das variáveis a) VBP_HA; b) GINI_AREA; c) CAD e d) resultado da regionalização.

4. Conclusões

O Coeficiente de Gini aplicado a concentração de terras por município e a Capacidade de Armazenamento de água Disponível no solo foram as variáveis que mais se correlacionaram com o Valor Bruto da Produção Agropecuária por unidade de área. As demais variáveis fisiográficas e sócio-econômicas não apresentaram relações com o Valor Bruto da Produção Agropecuária. Os planos de componentes indicam que as variáveis IDH_E, IDH_R, IDH_M e R_CAPTA são altamente correlacionadas, o que pode gerar problemas de multicolinearidade ao serem inseridas no mesmo modelo. A inclusão das variáveis de posição mostrou que o Valor Bruto da Produção Agropecuária apresenta variação no sentido Leste-Oeste. Houve a formação de sete regiões, que guardam relação com o uso e ocupação do solo.

4. Referências Bibliográficas

Andetta, G.M.A.C. **Valor Bruto da Produção Agropecuária Paranaense 1997 e 2004**. SEAB/DERAL/DEB 89 p. 2006

Biswanger, H. Elgin, M. Quais são as perspectivas para a reforma agrária? **Pesquisa e Planejamento Econômico**, Rio de Janeiro: v.19, p.01-18, 1989.

Câmara, G.; Souza, R.C.M.; Freitas, U.M.; Garrido, J. SPRING: Integrating remote sensing and GIS by object-oriented data modelling **Computers & Graphics**, 20: (3) 395-403, May-Jun 1996.

Cunha, M.S.; Chilante, C.A.T. Caracterização da agropecuária paranaense na década 90. **Revista Paranaense de Desenvolvimento**. Curitiba, n.101, p.3-16, 2001.

- Fuentes Llanillo, R.; Del Grossi, M.D.; Santos, F.O.; Munhos, P.D.; Guimarães, M.F. Regionalização da agricultura do Estado do Paraná, Brasil. **Ciência Rural**, v.36, n.1, p.120-127, 2006.
- Gasques, J. G.; Rezende, G. C.; Verde, C. M. V.; Conceição, J. C. P. R.; Carvalho, J. C. S.; Salerno, M. S. **Desempenho e crescimento do agronegócio no Brasil**. Brasília: IPEA, 2004. Disponível em: <<http://www.ipea.gov.br/TemasEspeciais/agronegocio.pdf>> Acesso em: 03 dez. 2006.
- Haykin, S. **Redes Neurais: Princípios e prática**. 2ª ed. Porto Alegre: Bookman, 2001. 893p.
- Hu, B. A note on calculating the Gini index. **Mathematics and Computers in Simulation** v.39 p.353-358, 1995.
- Instituto Brasileiro de Geografia e Estatística (IBGE). **Censo Agropecuário 1995-1996**. Disponível em: <<http://www.ibge.gov.br/>>. Acesso em: 20 out. 2006
- Instituto Paranaense de Desenvolvimento Econômico e Social (IPARDES). **Perfil dos municípios**. Disponível em: <<http://www.ipardes.gov.br/>> Acesso em: 20 out. 2006.
- 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.
- Maack, R...**Geografia Física do Estado do Paraná**. 2a. ed. Livr. José Olympio Ed., Rio de Janeiro, 450 pp, 1981.
- Matrix Laboratory (MATLAB) 7.0**. The MathWorks, Inc. 2004.
- Ministério da Agricultura – SECRETÁRIA NACIONAL DE PLANEJAMENTO AGRÍCOLA. **Aptidão agrícola das terras do Paraná**. Brasília, 1981. v.21, 141p.
- National Aeronautics e Space Administration (NASA). **GEOCOVERT** - Applied Sciences Directorate. Disponível em: <<https://zulu.ssc.nasa.gov/mrsid/>> Acesso em: 16 out. 2006.
- Rabus, B. M.; Eineder, A. R.; Bamler R. The shuttle radar topography mission- a new class of digital elevation models acquired by spaceborne radar, **Photogramm. Rem. Sens.**, v. 57, p. 241-262. 2003.
- Rogers, S. K.; Kabrisky, M. **An introduction to biological and artificial neural networks for pattern recognition** Washington:SPIE, 1991. 220p.
- Schneider, S.; Waquil, P.D. Caracterização socioeconômica dos municípios gaúchos e desigualdades regionais. **Revista de Economia e Sociologia Rural**, Brasília, v.39, n.3, p.117-142, 2001.
- Silva, M. A. S. **Mapas auto-organizáveis na análise exploratória de dados geoespaciais multivariados**. 2004-03-08. 117 p. (INPE-12434-TDI/996). Dissertação (Mestrado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2004.
- Vesanto, J.; Himberg, J.; Alhoniemi, E.; Parhankangas, J. **SOM Toolbox for Matlab 5**. Technical Report A57, Helsinki University of Technology, 2000. disponível em: <<http://www.cis.hut.fi/projects/somtoolbox/>> acesso em: 18 out. 2006..