

Proposição e avaliação de um indicador de homogeneidade territorial: estudo de caso dos Territórios de Identidade baianos

Marcos Aurélio Santos da Silva¹

Joel Alves de Oliveira²

Leonardo Nogueira Matos²

Márcia Helena Galina Dompieri¹

¹ Embrapa Tabuleiros Costeiros

Av. Beira Mar, nº 3250, CEP 49025-040, Aracaju – Sergipe – Brasil

{marcos.santos-silva, marcia.dompieri}@embrapa.br

² Universidade Federal de Sergipe - UFS

CEP 49.100-000 – Tel.: (79) 2105-6678 – São Cristóvão – SE – Brasil

{leonardo, joel.oliveira}@dcomp.ufs.br

Abstract. The stratification of municipalities in homogeneous areas facilitates the planning and management of space, especially in large areas, since it allows greater coherence in the adoption of public policies and economic incentives. This paper proposes a homogeneity index to evaluate the degree of similarity of territories previously determined by regional public policies. The index works in conjunction with clustering techniques such as TerraSOM and K-means. The proposed index has been applied in the evaluation of 27 Territories of Identity, state of Bahia, created in 2007 by the Superintendency of Economic and Social Studies of Bahia. The data were composed of a set of 45 socioeconomic and land use variables. The analytical evaluation of the proposed index showed that it is simple to implement and interpret, can be applied with any clustering technique, and can be used in different contexts. The results have shown that to the Territories of Identity results have shown similar results for both clustering techniques, but in the presence of outliers, the homogeneity index was less stable when calculated by the k-means technique. There is not statistically significant spatial autocorrelation for the homogeneity index.

Keywords: Indicator of homogeneity, spatial dependence, TerraSOM .

1. Introdução

A adoção de políticas de desenvolvimento baseadas na regionalização tem sido a estratégia governamental para promover o desenvolvimento econômico, diminuir a desigualdade social e promover a conservação ambiental (LEITE S.P. e WESZ JÚNIOR (2012),BOUERI R.; COSTA (2013)). Destacam-se nesse cenário, o Programa Nacional de Desenvolvimento Territorial Rural e o Programa Territórios da Cidadania do Ministério do Desenvolvimento Agrário, os Territórios de Planejamento do estado de Sergipe e os Territórios de Identidade do estado da Bahia. Nestes casos, cada Território corresponde a um conjunto de municípios vizinhos e contíguos que compartilham características socioeconômicas similares segundo critérios objetivos, como o nível de atividade econômica, ou subjetivos, como o grau de pertencimento ou identidade.

Embora a homogeneidade seja um conceito central das políticas com enfoque territorial no Brasil, ainda não foi definida uma métrica para o seu cálculo em territórios já definidos. Em geral, o uso de medidas de homogeneidade é aplicado antes da definição dos aglomerados espaciais, por exemplo, nos processos de zoneamento agroecológico, ambientais ou de risco climático (SILVA e SANTOS (2011)). Para o caso de territórios já definidos, faz-se necessária a definição de métodos de avaliação da homogeneidade para que seja possível o monitoramento

e realinhamento das políticas de desenvolvimento territoriais, quando necessário. Uma primeira estratégia seria escolher um conjunto de variáveis aleatórias multidimensionais e avaliar, para cada Território, o nível de variação intra e inter Territórios. No entanto, essa estratégia assume que os Territórios foram definidos corretamente. Outra estratégia é particionar o conjunto multivariado de dados para somente depois verificar se há dependência espacial para as variáveis selecionadas.

Neste trabalho foi adotada a segunda estratégia e definido um índice de homogeneidade que avalia um território a partir do nível de agregação de suas unidades constituintes. O nível de agregação entre as unidades (municípios, no caso das políticas públicas citadas), pode ser determinado por qualquer algoritmo de análise de agrupamentos. Optou-se por avaliar a implementação do TerraSOM devido ao seu histórico positivo de análise de agrupamentos por área, (SILVA, SIQUEIRA e TEIXEIRA (2010), SILVA et al. (2011), SILVA et al. (2015)), e o método k -médias por ser um algoritmo largamente empregado na área. O índice de homogeneidade foi avaliado na comparação dos 27 Territórios de Identidade do estado da Bahia, criados em 2007 pela Superintendência de Estudos Econômicos e Sociais da Bahia (SEI), vinculada à Secretaria de Planejamento do Estado da Bahia (SEPLAN).

2. Materiais e Métodos

2.1. Dados

Os dados são compostos por um conjunto de 45 variáveis socioeconômicas e de uso da terra que apresentam aspectos relacionados a indicadores do Atlas de Desenvolvimento Humano, PNUD¹, do Bolsa Família e do Censo Agropecuário, IBGE², divididas em 7 grupos: atlas de desenvolvimento humano, bolsa família, condição do produtor, uso da terra, efetivo de bovinos, efetivo de caprinos e efetivo de ovinos. Foram analisados dois conjuntos de dados: o primeiro com todos os municípios e Territórios de Identidade, ou seja, 417 municípios e 27 Territórios; e o segundo sem as observações atípicas, isto é, observações com valores, para determinadas variáveis, distantes cinco desvios-padrão em relação à média. Para o segundo conjunto de dados, analisou-se 375 municípios e 26 Territórios, o Território Metropolitano de Salvador foi excluído porque continha muitas observações com variáveis apresentando valores atípicos.

2.2. Mapas auto-organizáveis e o sistema TerraSOM

Os Mapas Auto Organizáveis são um tipo de rede neural artificial com aprendizagem não-supervisionada que apresenta a propriedade de mapear ordenadamente os dados na grade de neurônios. Esse mapeamento preserva as propriedades estatísticas dos dados de entrada e permite que essa grade seja explorada para visualização e exploração dos dados. O processo de aprendizagem compreende três etapas: competitiva, quando cada vetor de entrada é associado a um neurônio, o BMU (do inglês, Best Matching Unit); cooperativa, quando é definida a vizinhança entre os neurônios; adaptativa, quando os vetores de peso são atualizados na direção dos vetores de entrada na intensidade inversa da distância entre o neurônio e o BMU de cada vetor de entrada KOHONEN (2001).

O método TerraSOM³ particiona os dados geoespaciais a partir da rede neural artificial SOM em três etapas, (SILVA et al., 2015). Na primeira, a rede neural é treinada a partir do conjunto de dados multivariados agregados por município, usando a distância euclidiana como

¹Atlas do desenvolvimento humano. Disponível em <http://www.atlasbrasil.org.br/>.

²Sistema IBGE de Recuperação Eletrônica (SIDRA). Disponível em <http://www.sidra.ibge.gov.br/>.

³O sistema TerraSOM é uma implementação, na forma de plugin, do método TerraSOM por meio da biblioteca de classes SOMCode que encapsula algoritmos relativos ao Mapa Auto-Organizável de Kohonen usados pelo método.

medida de similaridade entre as observações. Na segunda, a rede neural é automaticamente particionada em c grupos usando o algoritmo Costa-Netto (COSTA J. A. F.; ANDRADE-NETTO (2001)). Como cada município estará associado a um único neurônio artificial, tem-se que a partição da rede neural determina a divisão dos dados em c grupos. Na Terceira etapa o particionamento é avaliado por meio de dois índices, o Davies-Bouldin que avalia a distância entre os centróides de cada grupo (DAVIES D. L.; BOULDIN (1979)) e o Composed Density between and within clusters (CDBw) que avalia a densidade intra e inter grupos por meio de vetores de referência dos dados de entrada (HALKIDI M.; VAZIRGIANNIS (2002), WU S.; CHOW (2004), HALKIDI M.; VAZIRGIANNIS (2008)). O uso dos índices de validação permite escolher as configurações de rede que geraram os melhores resultados.

Os testes com o método TerraSOM foram conduzidos para 80 redes com arquiteturas distintas. Todas possuíam grade bidimensional hexagonal e foram treinadas com aprendizagem sequencial. Entretanto diferiam no número de linhas e colunas, raio inicial do processo de atualização dos vetores de código e no número máximo de iterações.

2.3. Indicador de homogeneidade

Uma forma de avaliar a homogeneidade de um Território já constituído a partir de um conjunto de variáveis aleatórias é aplicar um algoritmo de análise de agrupamentos (com ou sem restrição de contiguidade espacial) sobre o conjunto de dados e compará-la aos Territórios de Identidade. Quanto mais próximo for o alinhamento entre um grupo e um Território maior serão as chances do mesmo ser considerado homogêneo.

Para comparar os resultados entre a partição dos dados e o Território foi idealizado um indicador de homogeneidade que leva em consideração a área coberta, S_{max} , pelo grupo dominante por Território de Identidade i , além do número m_i de grupos encontrados nesse mesmo Território. Logo, dado um particionamento dos objetos geográficos (municípios) em c grupos a partir dos métodos k-médias e TerraSOM, define-se o índice de homogeneidade para cada Território de Identidade T_i da forma que segue.

Seja $\{S_m\}_i$ o conjunto do somatório das áreas de T_i associados a um conjunto de m_i grupos, $S_i = \sum_{j=1}^{m_i} S_j$ o somatório de todos os elementos de $\{S_m\}_i$ e $S_{max,i}$ o valor máximo desse mesmo conjunto. O índice de homogeneidade x_i para cada T_i será dado por:

$$x_i = \frac{S_{max,i}}{S_i + \frac{m_i-1}{c}(S_i - S_{max,i})} \quad (1)$$

O índice x_i assume valor máximo quando $S_{max,i}$ se iguala a S_i , ou seja, quando o Território de Identidade está associado a um único grupo, e valor mínimo quando m_i se iguala a c e $S_{max,i}$ se aproxima de $\frac{S}{m_i}$, portanto, quando todos os grupos estão representados no Território e o maior grupo se aproxima de $\frac{S}{m_i}$. Escalonando x_i para o intervalo $[0,1]$ tem-se o valor y_i dado por:

$$y_i = 1 + \frac{x_i - 1}{c}(c^2 + (c - 1)^2) \quad (2)$$

como cada método gerará mais de uma partição dos dados, optou-se por considerar a média \bar{y}_i de todas as partições consideradas válidas como o indicador de homogeneidade. A autocorrelação espacial do indicador foi avaliada por meio dos índices I de Moran e c de Geary (CLIFF A. D.; ORD (1973), BAILEY T. C.; GATRELL (1995)).

Observa-se que o indicador de homogeneidade é independente do algoritmo usado para particionamento de dados assim como do tipo de conjunto de dados multivariados usados para a análise de homogeneidade.

2.4. Programas Usados

Foi utilizado a linguagem R, versão 3.2.1, com os pacotes *mapprools* versão 0.8 e *spdep* versão 0.5 para manipulação de dados geográficos e cálculo dos índices de autocorrelação espacial (BIVAND R. S.; PEBESMA (2008)). Os mapas foram gerados a partir do programa *QGis* versão 2.6.0. A análise de agrupamento *k*-médias foi feita com base na implementação existente no pacote *stat* da linguagem R. A execução do método TerraSOM foi realizada também por meio de uma chamada às classes da biblioteca SOMCode, implementados em linguagem C++, a partir de um *script* em R.

3. Resultados

3.1. Análise de Agrupamentos

Observa-se, no gráfico A e B da Figura 1, que à medida que o número de agrupamentos cresce, o erro de quantização vetorial diminui nos dois conjuntos de dados. De fato, o número de grupos, *c*, depende muito do número total de neurônios, *m*, quanto maior *m*, maior será *c*. Os gráficos C e D da Figura 1 mostram os valores dos índices de validação de agrupamentos Davies-Bouldin (DB) e CD_{bw}.

O índice Davies-Bouldin identificou como sendo os melhores experimentos, índices com valores baixos, aqueles que retornaram um número de agrupamento acima de 64 para o conjunto completo de dados, e acima de 45 para o conjunto sem dados atípicos, nesse intervalo o valor do índice permaneceu abaixo de 0,2. Já o índice CD_{bw} apontou como os melhores experimentos, índices com valores elevados, aqueles com menor número de agrupamentos. Logo, para a análise posterior, na avaliação do índice de homogeneidade territorial, foram considerados as redes que apresentaram valores intermediários para os índices Davies-Bouldin e CD_{bw}.

Para o teste com o método *k*-médias foram avaliados 22 agrupamentos, com *k* variando entre 10 e 31.

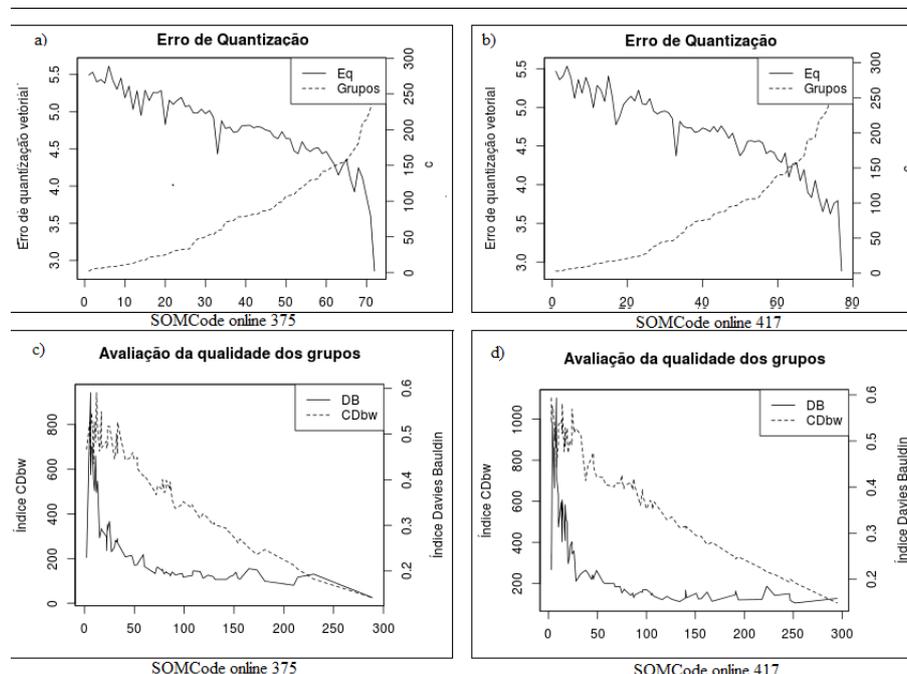


Figura 1: Gráficos da curva dos erros de quantização Eq, número de agrupamentos *c* para cada experimento em ordem crescente.

Pôde-se observar por meio da Figura 2, que ambas as técnicas tiveram resultados

semelhantes para os índices de homogeneidade por Território de Identidade quando consideradas as duas bases de dados, no entanto com maior variação para os índices calculados para a partição pelo método *k*-médias. Apesar de não coincidirem perfeitamente, os métodos *k*-médias e a implementação no TerraSOM geraram índices que classificaram os Territórios de forma bem aproximada.

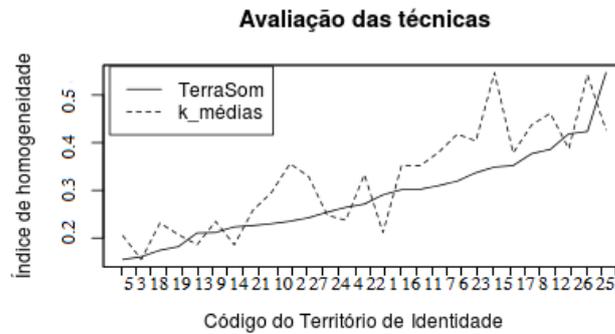


Figura 2: Curva da média dos índices de homogeneidade y_i por Território de Identidade

3.2. Índice de Homogeneidade

As Figuras 3 e 4 mostram a distribuição espacial dos índices de homogeneidade para os Territórios de Identidade com os dois conjuntos de dados, conjunto de dados completo (mapa A) e sem dados atípicos (mapa B). Os mapas coropléticos mostram os índices por meio das quebras claras, confirma-se por eles que há maior variação dos índices para a partição dos dados pelo método *k*-médias, e melhor separação dos dados para ambos os casos quando se elimina os dados atípicos. O cálculo dos índices *I* de Moran e *c* de Geary para os índices de homogeneidade não confirmaram autocorrelação espacial.

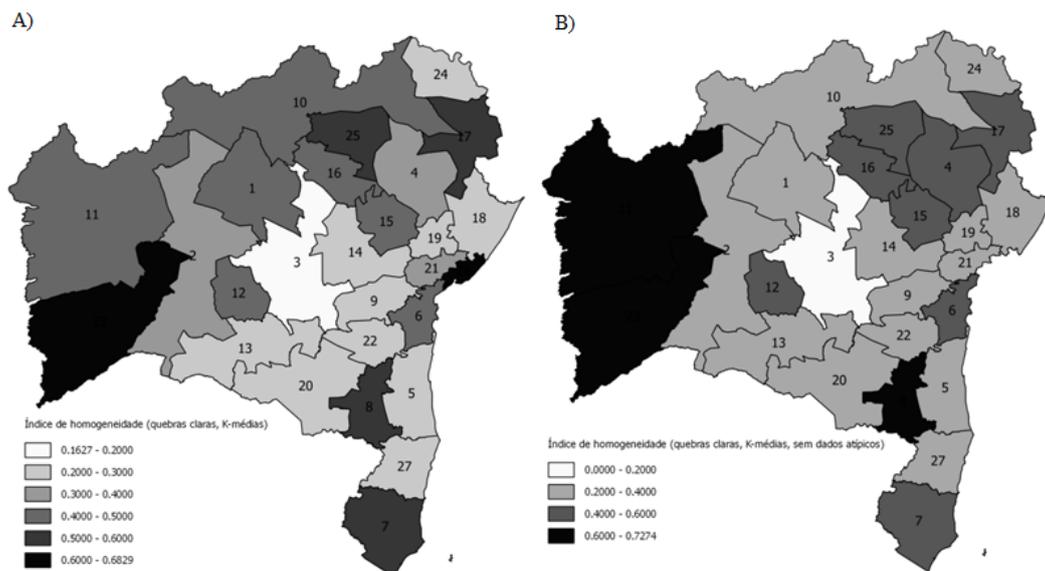


Figura 3: Mapas coropléticos para os índices de homogeneidade com particionamento por *k*-médias.

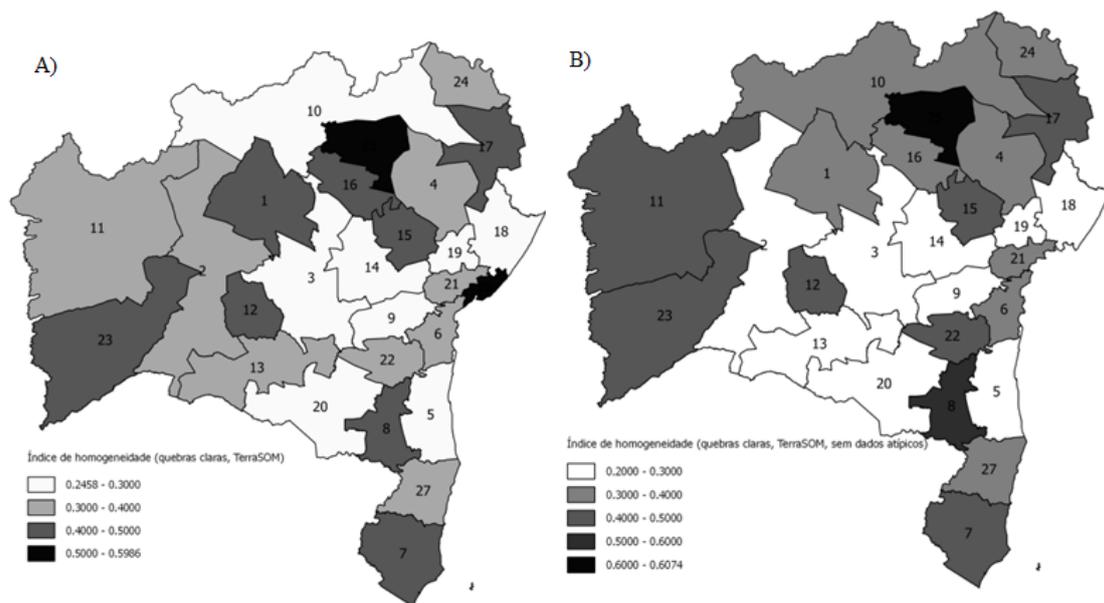


Figura 4: Mapas coropléticos para os índices de homogeneidade com particionamento por TerraSOM.

3.3. Análise do Índice de Homogeneidade

O índice proposto é simples, pode ser calculado a partir de qualquer algoritmo de análise de agrupamentos, é aplicável em diversos contextos onde se queira determinar o grau de homogeneidade em zonas ou territórios já demarcados e é de fácil interpretação. No entanto, a correspondência entre a homogeneidade calculada e efetiva deverá ser verificada por meio de trabalho de campo específico para os problemas da pesquisa ou das ações públicas desejadas. Outro fator relevante é a escolha das variáveis usadas na análise de agrupamentos. As mesmas podem ou não ter correspondência com os critérios usados para a criação dos territórios, pois o pesquisador pode querer verificar se um território criado a partir de determinados fatores também seria homogêneo a partir de outros.

4. Conclusões

O índice de homogeneidade calculado para os territórios baianos mostrou-se mais sensível à presença de dados atípicos quando calculado a partir do método k-médias, quando comparado ao método TerraSOM.

Não se observou autocorrelação espacial estatisticamente significativa para o índice de homogeneidade. Ao considerar o método TerraSOM, tem-se que foram identificados com clareza os Territórios de Identidades localizados nos extremos do território baiano.

O índice de homogeneidade proposto é de fácil aplicação e interpretação, sendo necessários estudos complementares tanto para analisar a sua resposta para outros algoritmos de análise de agrupamentos (e.g., que considerem questões de vizinhança espacial como *SKATER*), quanto para verificar a sua utilidade em diferentes problemas de análise de homogeneidade em partições pré-definidas (e.g., análise de homogeneidade de áreas de zoneamentos agroecológicos a partir de variáveis socioeconômicas).

Referências

BAILEY T. C.; GATRELL, A. C. Interactive spatial data analysis. *Essex: Longman*, 1995.

- BIVAND R. S.; PEBESMA, E. J. G.-R. V. Applied spatial data analysis with r. *New York: Springer*, 2008.
- BOUERI R.; COSTA, M. A. Brasil em desenvolvimento 2013: Estado, planejamento e políticas públicas. *Brasília: Ipea*, 2013.
- CLIFF A. D.; ORD, J. K. Spatial autocorrelation. *London: Pion*, 1973.
- COSTA J. A. F.; ANDRADE-NETTO, M. L. Clustering of complex shaped data sets via kohonen maps and mathematical morphology. *Data Mining and Knowledge Discovery*, p. 16–27, 2001.
- DAVIES D. L.; BOULDIN, D. W. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 1, n. 2, p. 224–227, April 1979.
- HALKIDI M.; VAZIRGIANNIS, M. Clustering validity assessment using multi representatives. *Hellenic Conference on Artificial Intelligence*, v. 2, 2002.
- HALKIDI M.; VAZIRGIANNIS, M. A density-based cluster validity approach using multi-representatives. *Pattern Recognition Letters*, v. 29, n. 6, p. 773–786, 2008.
- KOHONEN, T. Self-organizing maps. *New York: Springer*, v. 3, 2001.
- LEITE S.P. E WESZ JÚNIOR, V. Um estudo sobre o financiamento da política de desenvolvimento territorial no meio rural brasileiro. *Revista de Economia e Sociologia Rural, Piracicaba-SP*, v. 50, n. 4, p. 645–666, out/dez 2012.
- SILVA, J.; SANTOS, R. Estratégia metodológica para zoneamento ambiental: a experiência aplicada na bacia hidrográfica do alto rio taquari. *Campinas/SP: Embrapa Informática Agropecuária*, 2011.
- SILVA, M. A. S. et al. TerraSOM: Sistema para análise de dados geoespaciais agregados por Área baseado na rede neural do tipo mapa auto-organizável de kohonen. *Aracaju: Embrapa Tabuleiros Costeiros. Boletim de Pesquisa e Desenvolvimento.*, 2015.
- SILVA, M. A. S. et al. Using self-organizing maps for rural territorial typology. In: PRADO, Hercules Antonio; LUIZ, Alfredo Jose Barreto; CHAIB FILHO, Homero (Org.). *Computational methods for agricultural research*, v. 7, p. 107–126, 2011.
- SILVA, M. A. S.; SIQUEIRA, E. R.; TEIXEIRA, O. Abordagem conexionista para análise espacial exploratória de dados socioeconômicos de territórios rurais. *Revista de Economia e Sociologia Rural*, v. 48, p. 429–446, 2010.
- WU S.; CHOW, T. W. Clustering of the self-organizing map using a clustering validity index based on inter-cluster and intra-cluster density. *Pattern Recognition*, v. 37, p. 175–188, 2004.