

## Observação dos efeitos do problema da área modificável (MAUP) em dados de incidência de dengue no município de Natal, Brasil.

Igor Peregrino da Silva Sena<sup>1</sup>, Luiz Antonio Nascimento de Paiva<sup>2</sup>

<sup>1</sup> Instituto Nacional de Pesquisas Espaciais  
Caixa Postal 515 - 12227-010 - São José dos Campos - SP, Brasil  
igorsena@dsr.inpe.br

<sup>2</sup> Universidade Federal do Rio Grande do Norte – UFRN  
Caixa Postal 1607 – 59078-970 – Natal – RN, Brasil  
lanpaiva@ig.com.br

**Abstract.** This paper discuss about the MAUP effects found when comparing dengue incidence data with census data. Dengue is the most important arbovirus impacting the humankind, thus being a main public health problem of the world. Many works discuss this issue, even using GIS to draw its conclusions. Once punctual data is very difficult to obtain or generate, aggregated or areal data is the most used in many geographical inferences. The use of variables for areal units for modeling the occurrence of diseases, mainly those of census, may lead to a problem when zoning the regions of interest. There are many ways to create zones and it's noticed that changes in the zoning of a same area may affect the geographical analysis, consisting in the modifiable areal unit problem (MAUP). This work aims to verify the occurrence of the MAUP when comparing dengue incidence data with census socioeconomic data of the city of Natal, Brazil. The data has been zoned with two methodologies for 2007, 2008 and 2009. It was observed the variation in the correlation coefficient between different zoning outputs, with growing number of groups or with different ways of making the same number of groups. One methodology consists in generate groups with nearly the same number of dengue cases, the other consists in generating groups using a minimum spanning tree, maximizing the similarity between dengue incidence and demographic density and fixing the output number of groups. The MAUP was detected with its scale problem and its aggregation problem.

**Palavras-chave:** SIG, regionalização, dados agregados, problema de escala, problema de agregação.

### 1. Introdução

A dengue é atualmente a mais importante arbovirose que afeta o ser humano, constituindo sério problema de saúde pública no mundo. As condições do meio ambiente tropical favorecem o desenvolvimento e a proliferação do *Aedes aegypti*, principal mosquito vetor (BRASIL, 2005). Dada a sua crescente importância, mais e mais trabalhos tratam deste assunto na atualidade e muitos deles se apóiam nos SIG's para tecer suas conclusões como é o caso de Nascimento e Petta (2011) e Petta e Nascimento (2009) para o estado do Rio Grande do Norte.

A utilização de variáveis agregadas, notadamente as censitárias, pode levar a um problema quando se é necessário regionalizar as áreas de interesse. Existem inúmeras maneiras de se gerar novos grupos de áreas e é importante notar que os efeitos dessa variação de grupos podem influenciar as análises geográficas (Openshaw, 1984). Este é o problema da área modificável. Logo, os resultados estatísticos possuem validade dependente da unidade de área e do reconhecimento dos problemas existentes nas análises feitas com base em dados agregados (Dias et al., 2002).

Procedimentos automáticos de regionalização oferecem metodologias mais eficientes, sistemáticas e objetivas para geração de agregações ótimas do que métodos manuais, embora seu sucesso ainda dependa do quanto é possível modelar o fenômeno do mundo real, se é válida a parametrização do critério necessário, e da eficiência do algoritmo de regionalização (Cockings et al., 2011).

O SKATER é um programa de regionalização que se baseia numa árvore geradora mínima onde as áreas são tratadas como grafos que possuem um custo nas arestas que ligam

seus centróides, sendo as arestas com menor custo retiradas do grafo de modo que seja atendido o número de grupos solicitado ao programa. Ele tenta aumentar a homogeneidade interna dos grupos e aumentar a heterogeneidade entre grupos (Assunção et al., 2004).

O AZTool é um programa de regionalização que tenta formar grupos homogêneos dado um valor médio a ser alcançado e um limiar mínimo e máximo a ser respeitado. O programa também pode utilizar variáveis de homogeneidade para garantir que os grupos atendam a uma métrica de similaridade entre variáveis. A partir dos dados de entrada, o programa faz alterações das áreas agregadas adicionando ou removendo áreas contíguas aos grupos tantas vezes quanto for solicitado (iterações) e refaz este processo um dado número de vezes (repetições), retornando o melhor resultado obtido.

A maior ocorrência de chuvas e aumento da temperatura no verão favorece um padrão sazonal de incidência da dengue coincidente com esta estação. Embora ela seja mais comum nos núcleos urbanos, a doença pode ocorrer em qualquer localidade desde que exista população humana susceptível, presença do vetor e o vírus seja introduzido (BRASIL, 2005). Segundo REDE (2002) a taxa de incidência de dengue pode ser calculada a partir do número de casos novos confirmados de dengue por 100.000 habitantes, na população residente em determinado espaço geográfico, no ano considerado.

Dada a estrutura de proliferação da dengue, espera-se que a maior densidade de habitantes por área seja um fator determinante na caracterização da incidência da doença. Outro fator levado em conta neste trabalho é que domicílios que não estejam ligados à rede geral de água tenham maior suscetibilidade de estocar água, aumentando a quantidade de locais que o mosquito possa ovopositar.

Com isto, este trabalho visa observar os efeitos do problema da área modificável sobre um conjunto de dados de incidência de dengue do município de Natal-RN em comparação com dados socioeconômicos. O município de Natal é um dos muitos municípios brasileiros que apresentam epidemias intermitentes de dengue. Sua localização pode ser vista na figura 1.

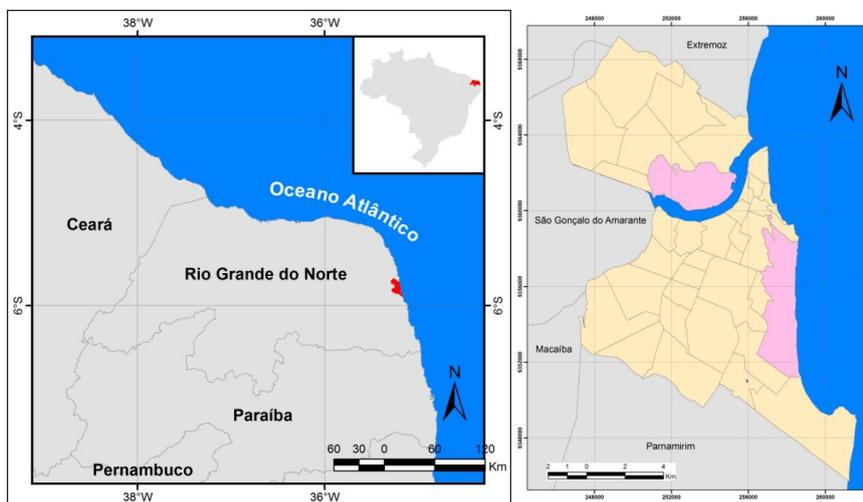


Figura 1 – Localização do município de Natal. No detalhe a direita vê-se a divisão dos bairros do município com destaque em magenta aos bairros de Salinas e Parque das Dunas.

## 2. Materiais e métodos

Este trabalho utiliza uma base de dados socioeconômicos por bairro do município de Natal-RN disponível nos anuários da cidade do Natal 2007, 2009 e 2010 (Natal, 2008; 2009; 2010) e uma base com dados de ocorrência de dengue espacializados por bairro do município de Natal em formato shapefile fornecido pelo LAGEOMA-UFRN. Os bairros de Salinas e Parque das Dunas foram excluídos das análises devido a sua grande extensão e baixo número de habitantes.

Os casos de dengue foram transformados em indicador pela seguinte fórmula:

$$\text{Indicador dengue} = \frac{\text{número de casos}}{\text{População}} * 100000 \quad (1)$$

Os dados socioeconômicos de área, número de domicílios (2007), população (2007), estimativa de população (2008 e 2009), ligações de água (2007, 2008 e 2009) foram integrados à tabela de atributos do shapefile.

Dois índices foram calculados neste trabalho. Um consiste na densidade populacional (hab/m<sup>2</sup>), sendo calculado pelo número de habitantes dividido pela área do bairro. O segundo é a porcentagem de domicílios sem ligação de água no bairro:

$$\% \text{ sem água} = \frac{\text{domicílios} - \text{ligações de água}}{\text{domicílios}} * 100 \quad (2)$$

Os 35 bairros da cidade passaram por dois procedimentos. O primeiro passo consistiu em regionalizar utilizando o software AZTool (Figura 2A), onde os grupos deveriam respeitar apenas a quantidade média de casos de dengue de 250, 500, 750, 1000, 1500, 2000 e até 3000 casos e desvios máximos de 50% para mais e menos. Nos casos em que algum bairro já possuísse o número de casos maior que o limiar máximo, o programa não consegue apresentar solução, sendo o limiar máximo afrouxado para o número de casos do bairro de maior ocorrência. O número de iterações utilizado foi de 100 e o número de repetições 150, onde é escolhida a melhor dentre as 150 soluções obtidas em sua execução. Em seguida os índices foram recalculados para cada grupo e foram computadas as correlações das regressões lineares entre o índice de dengue e a densidade populacional e entre o índice de dengue e o número de domicílios sem ligação de água.

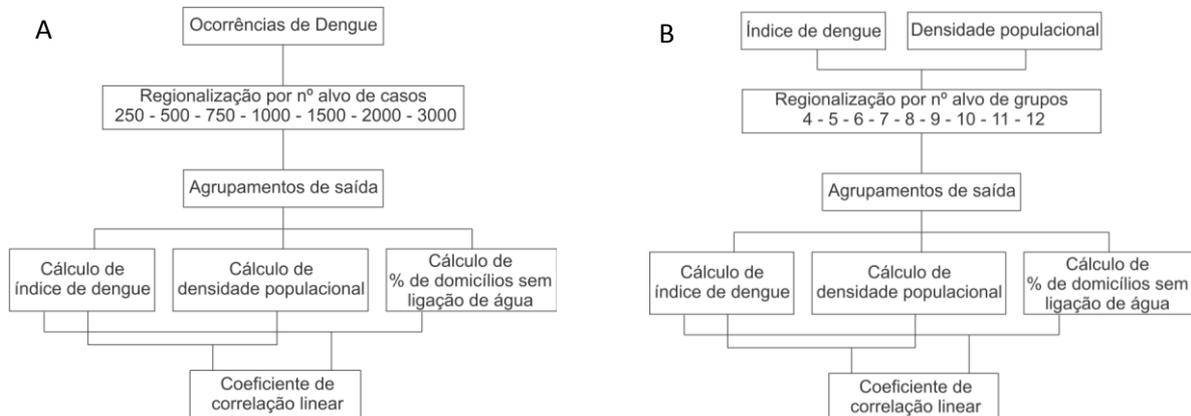


Figura 2. (A) Fluxograma da metodologia de regionalização AZTool. (B) Fluxograma da metodologia de regionalização SKATER.

O segundo passo consistiu em regionalizar os bairros através do software SKATER (Figura 2B), maximizando a similaridade entre o indicador da doença e a densidade populacional dentro de cada grupo para cada um dos anos estudados, gerando 4, 5, 6, 7, 8, 9, 10, 11 até 12 grupos. Em seguida os índices foram recalculados para cada grupo e foram computadas as correlações das regressões lineares entre o índice de dengue e a densidade populacional e entre o índice de dengue e o número de domicílios sem ligação de água.

### 3. Resultados e discussões

Nesta seção serão exibidos os resultados obtidos com os dois passos da metodologia.

### 3.1 Variações de escala

A metodologia AZTool gerou grupos com médias de ocorrência de dengue semelhantes e pré-determinadas em 250, 500, 750, 1000, 1500, 2000 e 3000 casos, ou até chegar em apenas 3 grandes grupos, devido ao uso de várias iterações e várias re-análises, mesmo o resultado sendo variável a cada utilização do programa, eles convergem para uma solução bem próxima. As correlações entre variáveis socioeconômicas obtidas através destes agrupamentos são exibidas nas tabelas 1, 2 e 3.

Tabela 1. Coeficientes de correlação linear entre o índice de dengue e a densidade populacional e entre o índice de dengue e a porcentagem de domicílios sem água obtidas para cada resultado através da metodologia AZTool para o ano de 2007.

Correlações AZTool 2007			
Valor alvo	Número de grupos	Densidade Populacional	Domicílios sem ligação de água
250	17	0,5203	-0,4919
500	11	0,8649	-0,7654
750	8	0,5208	0,2170
1000	6	0,8171	0,2932
1500	4	-0,0131	-0,8660

Tabela 2. Coeficientes de correlação linear entre o índice de dengue e a densidade populacional e entre o índice de dengue e a porcentagem de domicílios sem água obtidas para cada resultado através da metodologia AZTool para o ano de 2008.

Correlações AZTool 2008			
Valor alvo	Número de grupos	Densidade Populacional	Domicílios sem ligação de água
250	20	0,5787	-0,0114
500	18	0,5613	-0,6314
750	16	0,6205	-0,6352
1000	13	0,7453	-0,3087
1500	10	0,8842	-0,3134
2000	8	0,5132	-0,6028
3000	5	0,6897	0,7432

Tabela 3. Coeficientes de correlação linear entre o índice de dengue e a densidade populacional e entre o índice de dengue e a porcentagem de domicílios sem água obtidas para cada resultado através da metodologia AZTool para o ano de 2009.

Correlações AZTool 2009			
Valor alvo	Número de grupos	Densidade Populacional	Domicílios sem ligação de água
250	6	0,3729	-0,0583
500	3	0,7659	-0,9978

Observando os valores de correlação obtidos dos agrupamentos pela metodologia AZTool vê-se que os mesmos não aparentam seguir um padrão crescente ou decrescente para nenhuma das duas variáveis de comparação. Como o software trabalhou sem a utilização de uma

medida de homogeneidade, os grupos formados não necessariamente possuem a melhor correlação com uma ou outra variável representando apenas um dos cenários de agregação possível dos dados. Porém estes dados possuem como peculiaridade a formação de grupos que possuem um valor médio de casos de dengue bem similar sendo parecidos em relação à doença, mas bem erráticos em relação as variáveis socioeconômicas avaliadas.

A metodologia SKATER foi utilizada para gerar agrupamentos com número crescente de grupos. Os coeficientes de correlação linear entre variáveis socioeconômicas obtidas através destes agrupamentos são exibidos na tabela 4 e nas figuras 4A e 4B.

Tabela 4. Coeficientes de correlação linear entre o índice de dengue e a densidade populacional e entre o índice de dengue e a porcentagem de domicílios sem água obtidas para cada resultado através da metodologia SKATER para os anos de 2007, 2008 e 2009.

Número de Grupos	Correlação 2007		Correlação 2008		Correlação 2009	
	Densidade Populacional	Domicílios sem ligação de água	Densidade Populacional	Domicílios sem ligação de água	Densidade Populacional	Domicílios sem ligação de água
12	-	-	0,6320	-0,6151	-	-
11	-	-	0,6408	-0,5845	0,5684	-0,6366
10	-	-	0,6648	-0,5688	0,5899	-0,6364
9	0,1438	0,0154	0,6482	-0,6814	0,6464	-0,6430
8	0,2322	0,0131	0,6963	-0,7200	0,6466	-0,7222
7	0,2165	0,0802	0,9005	-0,7145	0,6069	-0,7860
6	0,3313	-0,0422	0,9333	-0,6970	0,5238	-0,7190
5	0,5078	0,0277	0,9888	-0,7003	0,7399	-0,6557
4	0,9426	-0,8911	0,9794	-0,8894	0,9214	-0,6147

Observa-se na tabela 4 e na figura 4A que a correlação entre o índice de dengue e a densidade populacional, diferentemente dos resultados obtidos com a metodologia AZTool, aumenta com a redução do número de grupos. Este efeito já era esperado dado que com a redução do número de grupos tem-se uma suavização dos indicadores resultando numa maior correlação entre as variáveis. No nível de bairro, todas as variáveis possuem uma correlação muito fraca com o índice da dengue que, nesse caso, foi artificialmente elevada com a redução da escala de trabalho. Este efeito constitui claramente a variação de escala do problema da área modificável.

A figura 4A mostra um comportamento decrescente para os coeficientes de correlação linear que atingem um patamar a partir de nove grupos para o ano de 2008 e de seis grupos para o ano de 2009, a partir do qual eles variam pouco. É importante notar que mesmo assim, as correlações continuam mais elevadas que o valor inicial obtido ao nível de bairro. Como os agrupamentos foram modelados de acordo com a relação entre o índice de dengue e a densidade populacional, a figura 4B mostra que a variável porcentagem de domicílios sem ligação de água não possui um comportamento bem determinado nas várias escalas de análise a não ser no caso de 4 grupos onde os coeficientes de correlação linear foram muito altos. Desta forma, os valores de coeficientes de correlação obtidos não possuem um significado espacial interpretável para as metodologias abordadas neste trabalho, servindo apenas de maneira ilustrativa como a representação de um cenário possível para esta variável naquela determinada escala de análise.

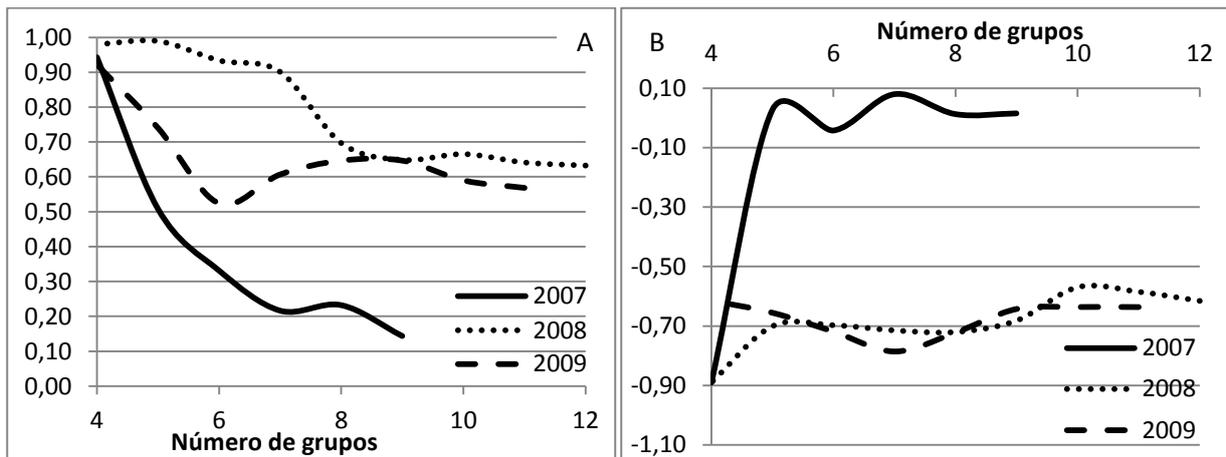


Figura 4. (A) Gráfico do coeficiente de correlação linear entre índice de dengue e a densidade populacional versus o número de grupos gerados. (B) Gráfico do coeficiente de correlação linear entre índice de dengue e o percentual de domicílios sem ligação de água versus o número de grupos gerados.

### 3.2 Variações de agregação

Os coeficientes de correlação obtidos com as duas metodologias foram comparados, quando possível, para avaliação das suas diferenças. Desta forma se analisa duas maneiras possíveis de se obter o mesmo número de grupos de análise, ou seja, duas representações da mesma escala de análise. As diferenças obtidas podem ser vistas nas tabelas 5 e 6 que comparam os valores obtidos no ano de 2007 e de 2008 com as duas metodologias.

Tabela 5. Comparação dos valores de correlação obtidos com a metodologia AZTool e SKATER para o ano de 2007

Metodologia AZTool 2007				Metodologia SKATER	
Valor alvo	Número de grupos	Densidade Populacional	Domicílios sem ligação de água	Densidade Populacional	Domicílios sem ligação de água
750	8	0,5208	0,2170	0,2322	0,0131
1000	6	0,8171	0,2932	0,3313	-0,0422
1500	4	-0,0131	-0,8660	0,9426	-0,8911

Tabela 6. Comparação dos valores de correlação obtidos com a metodologia AZTool e SKATER para o ano de 2008

Metodologia AZTool 2008				Metodologia SKATER	
Valor alvo	Número de grupos	Densidade Populacional	Domicílios sem ligação de água	Densidade Populacional	Domicílios sem ligação de água
1500	10	0,8842	-0,3134	0,6648	-0,5688
2000	8	0,5132	-0,6028	0,6963	-0,7200
3000	5	0,6897	0,7432	0,9888	-0,7003

Embora este trabalho apresente apenas dois cenários de diferentes agrupamentos com mesmo número de grupos (Figura 5), já é possível notar que mesmo se trabalhando numa mesma escala, pode-se obter diferentes relações entre 2 variáveis bastando para isso se modificar a maneira de agrupá-las. Este sem dúvida representa o segundo efeito do problema da área modificável, o de escala.

É notável que a metodologia SKATER que busca a melhor agregação possível, obteve um cenário com menores valores de correlação para seis grupos em 2007 e dez grupos em 2008, sendo superada pela metodologia AZTool que gerou um cenário sem maximizar as relações entre incidência de dengue e a densidade populacional. Comparando estes dois cenários obtidos, observa-se que a árvore geradora mínima não pode modelar todos os cenários ideais dada a sua estrutura de regionalização. Sendo assim, é provável que a utilização da metodologia iterativa de mudança de grupos respeitando a contigüidade e com restrição por critérios de homogeneidade possa levar a melhores resultados. O grande problema é que para gerar tais índices de homogeneidade é preciso conhecer o grupo de saída esperado, pois é dele que deriva este índice.

#### 4. Conclusões

Os resultados obtidos com a utilização das duas metodologias mostraram-se bem divergentes, devido à utilização das suas diferentes abordagens, uma usando um gerador iterativo de grupos com sucessivas trocas e avaliação da preservação dos parâmetros pré-estabelecidos para os grupos (AZTool) e o outro utilizando cortes baseados em medidas de similaridade entre os elementos iniciais não agrupados.

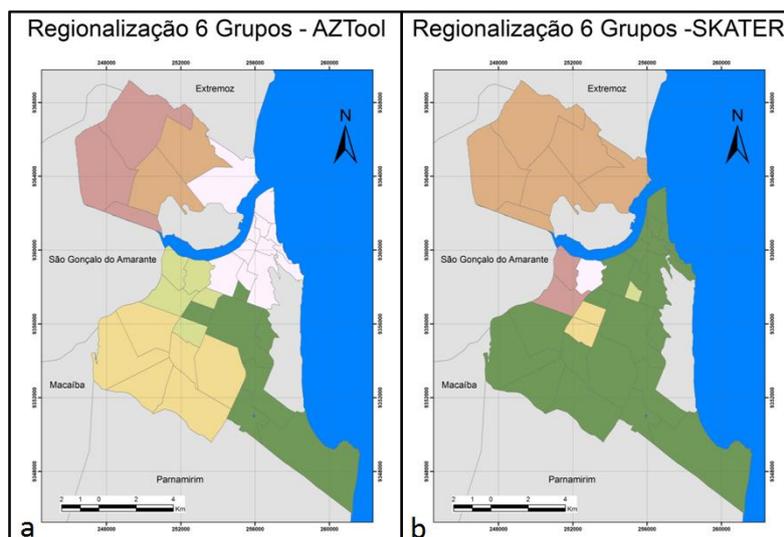


Figura 5. Dois cenários diferentes para a escala de agregação em seis grupos obtida com a metodologia AZTool (a) e a SKATER (b).

A solução por árvore geradora mínima (AGM) pode não encontrar o melhor cenário de regionalização possível com o máximo coeficiente de correlação linear entre duas variáveis, podendo o processo iterativo ser mais eficiente, principalmente quando da utilização de variáveis de homogeneidade. Uma pista desta possibilidade é o caso de 6 grupos da tabela 5 e figura 5 onde a agregação “aleatória” gerou um coeficiente de correlação linear maior que a correlação “ideal”. Contudo, mais testes são necessários para que se possa afirmar com certeza este fenômeno. Por outro lado, a utilização da AGM mostra bem o comportamento de *outliers*, áreas que não se encaixam bem em grupos e são isoladas das outras com o aumento do número de grupos. Esta pode ser uma ferramenta de identificação de regiões problema desde que se utilize variáveis com significado geográfico conhecido e relacionado ao problema em questão. Cabe ao usuário identificar sua necessidade e escolher a metodologia mais eficiente para a solução do seu problema.

Foram observados nos dados os dois efeitos relacionados ao problema da área modificável, escala e agregação, onde tanto diminuindo-se a escala, quanto variando-se a maneira de agregar as áreas, encontram-se diferentes relações entre duas variáveis. Tendo em

vista esse problema, é necessário que trabalhos que envolvam estas variáveis utilizem de cautela quando da afirmação da existência da dependência entre elas observando se não está ocorrendo apenas um dos dois efeitos acima citados. Este problema pode ser minimizado utilizando-se agregações em zonas com significado geográfico de onde possam se tecer relações reais entre variáveis.

Em suma, a análise de dados agregados por áreas deve ser cautelosa, e, como falado por outros autores, é de extrema importância se tomar conhecimento do problema da área modificável, e embasar suas discussões levando em conta a sua existência. Um meio prático de minimizar este problema é levar em conta aspectos geográficos relacionados ao seu problema e tornar este problema uma ferramenta a mais de análise.

### **Agradecimentos**

O autor deste trabalho gostaria de agradecer ao LAGEOMA-UFRN por fornecer os dados de ocorrência de dengue no município de Natal. Ao LESTE-UFMG pela disponibilização gratuita do software SKATER e pela pronta resposta de suporte do Msc. Elias Teixeira Krainski. Aos Drs. Samantha Cockings e David Martin e a universidade de Southampton que possuem os direitos e distribuem gratuitamente o software AZTool. A Dr<sup>a</sup> Liliam César de Castro Medeiros pela ajuda conceitual com o problema da dengue. E por fim, aos Drs. Antônio Miguel Vieira Monteiro, Eduardo Gerbi Camargo pelas orientações.

### **Referências**

- Assunção, R. M.; Lage J. P.; Reis, E. A. Análise de conglomerados espaciais via árvore geradora mínima. **Revista Brasileira de Estatística**, v. 49, n. 2, p. 177-196, 2004.
- BRASIL, Ministério da Saúde. **Guia de Vigilância Epidemiológica**. Brasília: Ministério da Saúde/ Secretaria de Vigilância em Saúde, 2005. 816 p.
- Cockings, S.; Harfoot, A.; Martin, D.; Hornby, D. Maintaining existing zoning systems using automated zone design techniques: methods for creating the 2011 census output geographies for England and Wales. **Environment and Planning A**, v. 43, p. 2399-2418, 2011.
- Dias, T. L.; Oliveira, M. P. G.; Câmara, G.; Carvalho, M.S. Problemas de escala e a relação área-indivíduo em análise espacial de dados censitários. **Informática Pública**, v. 01, n. 04, p. 89-104, 2002.
- Nascimento, P. S. R.; Petta, R. A. Uso de sistema de informação geográfica na dispersão de casos de dengue no estado do Rio Grande do Norte. In: Simpósio Brasileiro de Sensoriamento Remoto (SBSR), 15, 2011, **Anais...** São José dos Campos: INPE, 2011. Artigos, p. 8421-8428.
- Natal. Prefeitura Municipal do Natal. **Anuário Natal 2007**. Natal: Secretaria Municipal de Meio Ambiente e Urbanismo, 2008. 342 p.
- Natal. Prefeitura Municipal do Natal. **Anuário Natal 2009**. Natal: Secretaria Municipal de Meio Ambiente e Urbanismo, 2009. 402 p.
- Natal. Prefeitura Municipal do Natal. **Anuário Natal 2010**. Natal: Secretaria Municipal de Meio Ambiente e Urbanismo, 2010. 402 p.
- Openshaw, S. The modifiable areal unit problem. **Concepts and Techniques in Modern Geography**, n.38, p. 3-41.
- Petta, R. A.; Nascimento, P. S. R. Uso de sistema de informação geográfica na dispersão de casos de dengue entre 2000 e 2004 no estado do Rio Grande do Norte. **Geonordeste (UFS)**, v. 2, p. 207-226, 2009.
- REDE Interagencial de Informações para a Saúde. **Indicadores básicos de saúde no Brasil: conceitos e aplicações**. Brasília: Organização Pan-Americana de Saúde, 2002. 209 p.