

ÁRVORES DE DECISÃO APLICADAS À DEFINIÇÃO DE REGIÕES URBANAS HOMOGENEAS INCORPORANDO DADOS DE OFERTA DE INFRAESTRUTURA RODOVIÁRIA À GRADE ESTATÍSTICA

Caio Cesar dos Anjos Carneiro

Gustavo Garcia Manzato

Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP)

Faculdade de Engenharia de Bauru

RESUMO

Este trabalho apresenta um método para a definição de Regiões Urbanas Homogêneas (RUHs) com base em árvores de decisão, combinando-se dados de oferta de infraestrutura rodoviária com a grade estatística do IBGE (Instituto Brasileiro de Geografia e Estatística). Primeiramente, a oferta de infraestrutura rodoviária foi incorporada à cada célula da grade por meio de um indicador de cobertura espacial que contempla faixas de influência ponderadas ao redor das vias. Em seguida, a técnica ESDA (do inglês, *Exploratory Spatial Data Analysis*) foi aplicada para esse indicador resultando em uma classificação de cada célula segundo os quatro quadrantes do gráfico de Moran. Sobrepondo-se a base dos limites administrativos municipais à grade, foi calculada a proporção, em termos de área territorial para cada município, das células classificadas em cada quadrante e a área total do município (%HH, %LL, %LH e %HL). Tais proporções, juntamente com a porcentagem da área urbanizada sobre a área total de cada município (%AU), foram inseridas em um modelo de árvores de decisão, gerando regras que permitiram compor o método aqui desenvolvido. Dois modelos se destacaram, sendo o primeiro baseado apenas em %HH e o segundo apenas em %LL, mas ambos incluindo %AU.

ABSTRACT

This work presents a method for the definition of Functional Urban Regions (FURs) based on decision trees, combining road infrastructure supply data with the statistical grid from IBGE (Brazilian Institute of Geography and Statistics). Firstly, road infrastructure supply was incorporated into each grid cell by means of an indicator of spatial coverage that includes weighted influence buffers around road segments. Then, the Exploratory Spatial Data Analysis (ESDA) technique was applied to this indicator resulting in a classification of each cell according to the four quadrants of the Moran's scatterplot. Overlaying the municipal administrative boundaries database to the grid, we calculated the proportion, in terms of territorial area for each municipality, of cells classified in each quadrant and the total area of the municipality (%HH, %LL, %LH and %HL). These proportions, along with the percentage of the urbanized area over the total area of each municipality (%AU), were inserted into a decision tree model, generating rules that composed the method developed here. Two models were highlighted, the first being based only on %HH and the second only on %LL, but both including %AU.

1. INTRODUÇÃO

A definição de Regiões Urbanas Homogêneas (RUHs), termo empregado como referência geral às regiões metropolitanas, aglomerações urbanas etc., ainda é um desafio para os planejadores e gestores urbanos. Tal desafio reside no fato de que os limites dessas áreas urbanizadas normalmente diferem dos limites administrativos oficiais (os municípios, por exemplo) onde estão inseridas, e podem também abranger mais de um município vizinho.

Além dessa questão territorial, a metropolização gera inter-relações econômicas, sociais, políticas e demográficas entre os municípios que compõem essas RUHs, desencadeando processos urbanos que necessitam ser analisados em uma escala intermediária entre, por exemplo, os municípios e o estado. Nesse sentido, a figura de um governo metropolitano é geralmente constituída para esse fim. Entretanto, diversas abordagens podem ser encontradas na literatura objetivando a definição dessas RUHs, cada uma contando com suas particularidades em termos de resolução espacial, dados empregados e métodos aplicados.

Exemplos dessas abordagens incluem relações político-administrativas (São Paulo, 1989; Gouvêa, 2005; Arellano Ríos, 2014a, 2014b, 2015); recursos de sensoriamento remoto (Weber, 2001;

Huang et al., 2014; Niemeyer et al., 2014; Sahoo e Pekkat, 2014); aspectos econômicos, como o preço da terra (Bode, 2008) e a distribuição de empregos (Coombes e Overman, 2004); métricas espaciais para a determinação da forma urbana e sua morfologia (Bereitschaft e Debbage, 2014); redes de infraestrutura e serviços de transporte (Boarnet e Haughwout, 2000; Manley, 2014; Manzato et al., 2015; Guzik et al., 2017; Moisés et al., 2017); análise de movimentos pendulares (Cheshire e Hay, 1989; Konjar et al., 2010; Office of Management and Budget, 2010; Kauffmann, 2012; Williams et al., 2012; Horňák e Kraft, 2015; Klapka e Halás, 2016; OECD, 2016; Aguiar e Manzato, 2017; Aguiar et al., 2017; Kraft e Marada, 2017; Soares et al., 2017); distribuição territorial da população (Ramos e Rodrigues da Silva, 2003 e 2007; Ramos et al., 2004; Manzato et al., 2007; Santos et al., 2014; Souza e Manzato, 2017); combinações de indicadores populacionais e de oferta de infraestrutura de transportes (Manzato e Rodrigues da Silva, 2010; Pereira e Rodrigues da Silva, 2010; Ajauskas et al., 2012; Rodrigues da Silva et al., 2014; Matioli et al., 2017).

Se por um lado há diversas metodologias para a definição de RUHs, Kourtit et al. (2015) argumentam que, muitas vezes, a reprodução de certas abordagens é comprometida devido a definições ambíguas ou imprecisas, falta de dados específicos e/ou dificuldade de mensurar ou obter tais dados. Nesses casos, a distribuição populacional poderia ser empregada para se medir o grau de urbanização (UN, 2007; Office of Management and Budget, 1998), dada a relativa facilidade de obtenção de dados do censo populacional em praticamente todos os países, além desses dados serem bastante confiáveis. Ramos e Rodrigues da Silva (2003 e 2007), Ramos et al. (2004), Manzato et al. (2007), Santos et al. (2014) e Souza e Manzato (2017) exploraram essa hipótese com o emprego de ferramentas de estatística espacial, obtendo resultados positivos.

Em contrapartida, observa-se que processo de metropolização também está relacionado com o fluxo crescente de bens, serviços e oportunidades provenientes de demandas da população. Para isso, existe a necessidade de redes de infraestrutura, sobretudo a de transportes, as quais desempenham um importante papel na integração e organização do espaço. Alguns estudos (Boarnet e Haughwout, 2000; Manley, 2014; Manzato et al., 2015; Guzik et al., 2017; Moisés et al., 2017) procuraram empregar indicadores de oferta de infraestrutura de transportes objetivando a definição de RUHs. Especificamente, Manzato et al. (2015) mostraram que maiores níveis de oferta de infraestrutura rodoviária estão diretamente relacionados com a existência de RUHs.

Um dos problemas detectados na maioria dos estudos observados, entretanto, refere-se à questão da resolução espacial utilizada para o desenvolvimento das análises. Normalmente utiliza-se a subdivisão territorial dos municípios, agregando-se os dados nesta escala espacial. Embora as definições de RUHs devam necessariamente abranger o município como um todo (ou seja, não é possível tornar apenas parte do município pertencente a uma RUH), tal agregação espacial resulta no problema conhecido como “falácia ecológica”. Isso ocorre quando são realizadas análises com resultados derivados de agregação de valores por unidade de área, inferindo que estes valores correspondem ao nível individual (Jargowsky, 2005). Uma alternativa para contornar essa questão é o emprego e análise de dados em um nível mais desagregado primeiramente e, a partir dessa etapa, agregar os resultados na escala municipal, por exemplo.

Nesse sentido, Santos et al. (2014) e Souza e Manzato (2017) exploraram essa hipótese a partir da subdivisão territorial baseada nos setores censitários definidos pelo Instituto Brasileiro de Geografia e Estatística (IBGE). No entanto, esta escala espacial ainda não demonstrou resultados conclusivos. Por outro lado, o IBGE disponibilizou recentemente uma base georreferenciada com

dados do Censo de 2010, denominada Grade Estatística. Sua estrutura permite obter informações de maneira bastante detalhada em termos espaciais, uma vez que o território está dividido em células de 200 x 200 m nas áreas urbanas.

Com base nessas informações, o objetivo desse trabalho foi explorar recursos da grade estatística combinados com dados de oferta de infraestrutura rodoviária para a elaboração de um método de definição de RUHs. Tal método se baseia na ferramenta de árvores de decisão para a combinação e análise dos referidos dados visando a obtenção de regras de decisão. Foi utilizado um indicador de cobertura espacial que contempla faixas de influência ponderadas ao redor das vias, conforme Magalhães (2016) e a técnica ESDA (do inglês, *Exploratory Spatial Data Analysis*) aplicada para esse indicador resultando em uma classificação de cada célula da grade segundo os quatro quadrantes do gráfico de Moran. Essa técnica produz resultados bastante relevantes ao se comparar o valor de um atributo em um dado local com seus vizinhos, permitindo a caracterização da dependência espacial e análise do território de uma maneira integrada.

Na próxima seção do artigo será apresentada a metodologia desenvolvida. Na sequência, na seção 3 serão apresentados os resultados e discussões. Por fim, as conclusões do estudo podem ser encontradas na seção 4, seguidas da lista de referências citadas.

2. METODOLOGIA

Esta seção apresenta a proposta do método para a definição de RUHs com base na combinação de dados da grade estatística e da oferta de infraestrutura rodoviária. Primeiramente serão apresentadas algumas informações sobre as bases de dados utilizadas, seguidas da forma de obtenção do indicador de oferta de infraestrutura rodoviária e sua incorporação à grade. Na sequência, o emprego da técnica ESDA sobre o referido indicador será apresentado, resultando nas variáveis que serão inseridas no modelo de árvores de decisão, detalhadas ao final.

2.1. Bases de dados utilizadas e área de estudo

Três bases de dados principais foram utilizadas nesse estudo. A primeira refere-se à grade estatística do IBGE, a segunda contempla os limites administrativos dos municípios e a terceira corresponde à malha rodoviária. Tanto a base da grade como a base dos municípios podem ser obtidas junto ao IBGE, enquanto a base da malha rodoviária pode ser obtida junto ao Departamento de Estradas de Rodagem do Estado de São Paulo (DER-SP). Cabe destacar que o estudo de caso foi desenvolvido para o estado de São Paulo.

A grade estatística divide o território brasileiro em células de 200 x 200 m nas áreas urbanas e em células de 1 x 1 km nas áreas rurais, permitindo a recuperação de informações para diversos recortes espaciais independentes das unidades político-administrativas. Dessa maneira, considerou-se a sua utilização bastante pertinente ao proporcionar uma unidade de análise desagregada em termos espaciais. O principal dado extraído diretamente desta base foi a área urbana dos municípios, agregando-se as suas respectivas células de 200 x 200 m. Combinando-se esse dado com a base dos limites administrativos municipais, foi possível obter a primeira variável empregada no presente estudo, referente à porcentagem da área urbanizada sobre a área total de cada município (%AU).

A base da malha rodoviária foi utilizada para se determinar a oferta de infraestrutura rodoviária, computada para cada célula da grade. Um indicador conforme Magalhães (2016) foi empregado, que mensura a área influenciada por um sistema de transporte. A composição deste indicador,

mostrado na Equação 1, considera um conjunto de faixas (bandas ou *buffers*) ponderadas e equidistantes ao redor deste sistema, adotando-se o maior peso à faixa que está mais próxima do segmento rodoviário e o menor peso àquela que está mais distante.

$$IC_x = \frac{\sum_1^n \gamma(i) \cdot A_i}{A_x} \quad (1)$$

em que:

- IC_x : índice de cobertura espacial de estrutura de circulação viária para zona x (no caso, cada célula da grade estatística);
 $\gamma(i)$: função que determina o peso atribuído à i -ésima faixa de cobertura (*buffer*), tal que $\gamma(i) \in [0,1]$;
 A_i : área da i -ésima faixa contida em uma zona x ;
 A_x : área da zona x ;
 n : número de faixas (foram adotadas neste estudo 10 faixas com 1 km de equidistância cada).

2.2. Aplicação da técnica ESDA para o indicador de oferta de infraestrutura rodoviária

A técnica ESDA fundamenta-se na caracterização da dependência espacial, ou seja, indica como os valores estão correlacionados no espaço. De acordo com Anselin (1995 e 1998), essa técnica permite obter uma medida de similaridade (ou autocorrelação) espacial entre as áreas (ou zonas) de uma dada região, com relação a um determinado indicador. No caso deste estudo, trata-se do indicador de oferta de infraestrutura rodoviária apresentado na Equação 1.

A aplicação da técnica ESDA classifica cada célula da grade estatística por meio de dois parâmetros. O primeiro, Z , é função da diferença entre o valor do indicador de oferta de infraestrutura rodoviária em cada célula e a média global desse indicador considerando-se todas as células. O segundo, W_z , é função da diferença da média do referido indicador obtida nas células adjacentes a uma dada célula e a média global desse indicador considerando-se todas as células. As relações de adjacência são proporcionadas por uma matriz de proximidade espacial (W), que descreve o arranjo espacial das células da grade. É uma matriz quadrada, com $n \times n$ elementos, onde cada elemento, w_{ij} , representa uma medida de proximidade espacial entre a célula i e a célula j . Os parâmetros Z e W_z são usados para gerar os resultados (índices e classificações) que são utilizados em conjunto com as ferramentas de visualização de análise exploratória, como o gráfico de espalhamento de Moran, classificando os resultados em quatro quadrantes: HH, LL, LH e HL. Adicionalmente, essa classificação pode ser visualizada sobre o território em estudo por meio de um mapa temático, conhecido como *Box Maps*.

No quadrante HH (de *High-High*), o valor do indicador de oferta de infraestrutura rodoviária na célula é superior à média global e a média de seus vizinhos também é maior que a média global. No quadrante LL (de *Low-Low*), o valor do referido indicador na célula é inferior à média global e a média de seus vizinhos também é inferior à média global. Dessa forma, nesses dois quadrantes, visualiza-se uma autocorrelação espacial positiva. Já os quadrantes LH (de *Low-High*) e HL (de *High-Low*) apresentam uma autocorrelação espacial negativa, tendo em vista que a célula em análise difere de seus vizinhos em relação ao atributo estudado. No quadrante LH, o valor do indicador de oferta de infraestrutura rodoviária na célula é inferior à média global e a média dos seus vizinhos é superior à média global. E no quadrante HL, o valor do referido indicador na célula é superior à média global e a média de seus vizinhos é inferior à média global.

Uma vez determinadas as classificações de cada célula segundo os quadrantes gráfico de espalhamento de Moran para indicador de oferta de infraestrutura rodoviária, a base dos limites administrativos municipais foi sobreposta à grade, em que foi calculada a proporção, em termos de área territorial para cada município, das células classificadas em cada quadrante e a área total do município. Disso resultou nas outras variáveis empregadas nesse estudo: %HH, %LL, %LH e %HL. Ou seja, a variável %HH representa, por exemplo, a porcentagem da área das células classificadas como HH dentro de um município sobre a área total desse município.

Cabe ressaltar que todas as informações são obtidas primeiramente na escala espacial da grade estatística e posteriormente elas serão agregadas à escala dos municípios. Isso se deve ao fato de que as RUHs são definidas necessariamente para o município como um todo, ou seja, não é possível tornar apenas parte dele pertencente a uma RUH.

As etapas descritas nos itens 2.1 e 2.2 foram auxiliadas pelos *softwares* de Sistema de Informação Geográfica (SIG) QGIS versão 3.4 e de estatística espacial GeoDa versão 1.12, ambos de distribuição gratuita.

2.3. Aplicação das árvores de decisão para a definição de RUHs

As árvores de decisão constituem ferramentas de suporte à decisão baseadas em um fluxograma, cuja estrutura representa um modelo estatístico para a classificação e previsão de dados (Kamiński et al., 2017). Devido à sua semelhança com uma árvore, sua estrutura considera ramos e nós. O primeiro nó da árvore, chamado nó raiz, armazena todo o banco de dados. A partir nó raiz, com base na estrutura dos atributos observados no banco de dados e seus respectivos valores, a árvore se subdivide em outros nós, chamados de nós filhos.

Os ramos conectam o nó raiz com seus nós filhos e carregam em sua estrutura o resultado de uma decisão. Esse processo se repete até que não seja mais possível ocorrer uma nova subdivisão e, nesse caso, obtém-se o último nó, chamado de nó terminal ou nó folha. O caminho desde o nó raiz até cada nó folha é único e representa uma regra de classificação. O resultado é uma hierarquia de sentenças do tipo “se... então...” que são utilizadas para a classificação dos dados por meio de regras de decisão (Quinlan, 1987).

O banco de dados utilizado para desenvolver as análises com árvores de decisão constituiu a base dos 645 municípios do estado de São Paulo, em que foram incorporadas as variáveis descritas anteriormente, a saber: porcentagem da área urbanizada sobre a área total de cada município (%AU) e porcentagem da área das células classificadas em cada quadrante dentro de um município sobre a área total desse município (%HH, %LL, %LH e %HL). Essas corresponderam às variáveis independentes dos modelos. A variável dependente (RUH) se referiu a uma classificação binária para cada município sobre o fato dele já pertencer a uma RUH oficial ou não, ou seja, se sim, $RUH = 1$; se não, $RUH = 0$.

O *software* IBM SPSS Statistics versão 25 foi utilizado para a geração das árvores de decisão. Um estudo exploratório de uma série de modelos foi desenvolvido, analisando o conjunto de variáveis que produziam o modelo de melhor desempenho. O mesmo ocorreu em relação ao algoritmo de criação da árvore, em que foram explorados os métodos CHAID, *Exhaustive* CHAID, C&RT e QUEST.

3. RESULTADOS

A partir do que foi descrito nos itens 2.1 e 2.2, a Figura 1 apresenta um mapa temático contendo a grade estatística referente ao estado de São Paulo, cujas células estão classificadas segundo os quadrantes do gráfico de espalhamento de Moran para indicador de oferta de infraestrutura rodoviária. A base de dados das rodovias que deu origem ao indicador também está sobreposta à grade. No detalhe, é possível observar a estrutura espacial da grade, com as células de 200 x 200 m e as células de 1 x 1 km.

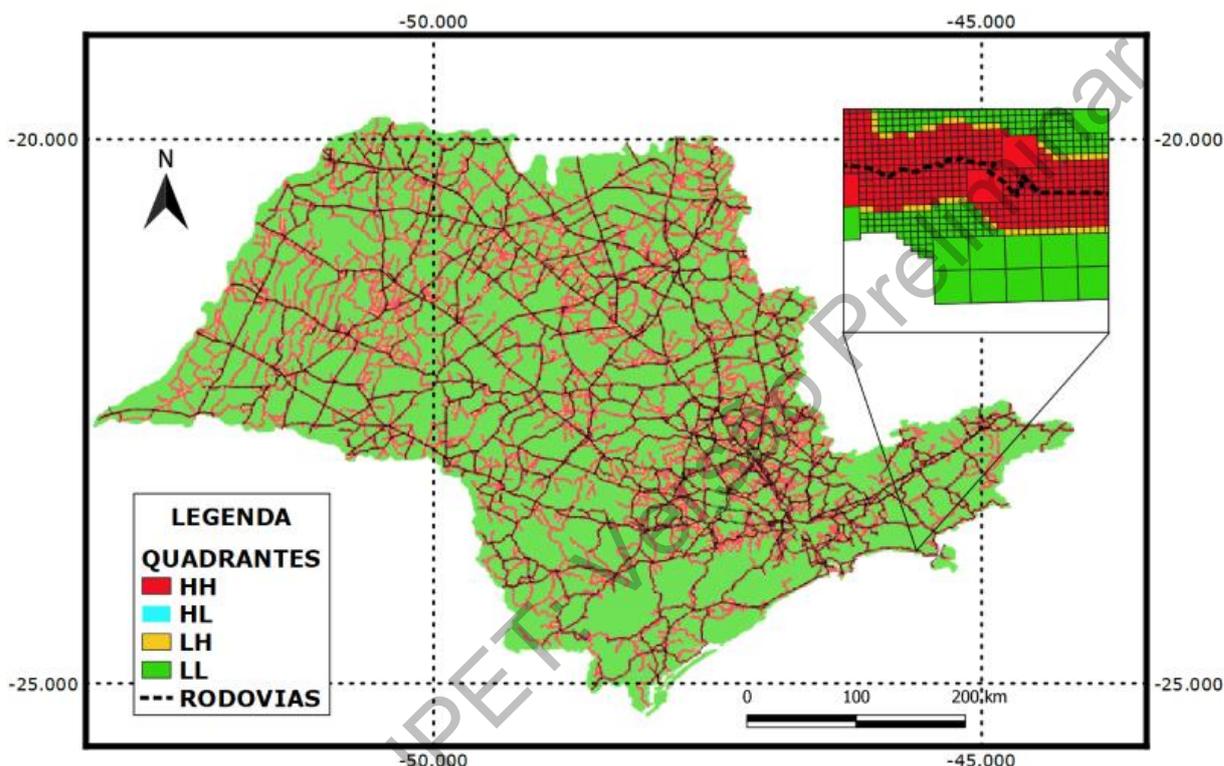


Figura 1: Mapa temático a grade estatística referente ao estado de São Paulo com as células classificadas segundo os quadrantes do gráfico de espalhamento de Moran para o indicador de oferta de infraestrutura rodoviária

Na sequência, a Figura 2 ilustra no mapa à esquerda a estrutura espacial da grade para a obtenção da variável %AU para um dado município. No mapa à direita, pode-se observar um detalhe da grade estatística com as células classificadas segundo os quadrantes do gráfico de espalhamento de Moran para a obtenção das variáveis %HH, %LL, %LH e %HL referentes ao indicador de oferta de infraestrutura rodoviária.

Em relação à série de modelos de árvores de decisão explorados, verificou-se que as árvores baseadas no algoritmo QUEST produziram os melhores desempenhos. Em termos das variáveis incluídas, a variável %AU mostrou-se relevante para todos os modelos analisados. Já as outras variáveis (isto é, %HH, %LL, %LH e %HL) apenas %HH e %LL apresentaram resultados consistentes, mas em modelos separados. Em resumo, dois modelos se destacaram, sendo o primeiro baseado apenas em %HH e em %AU, e o segundo baseado apenas em %LL e em %AU.

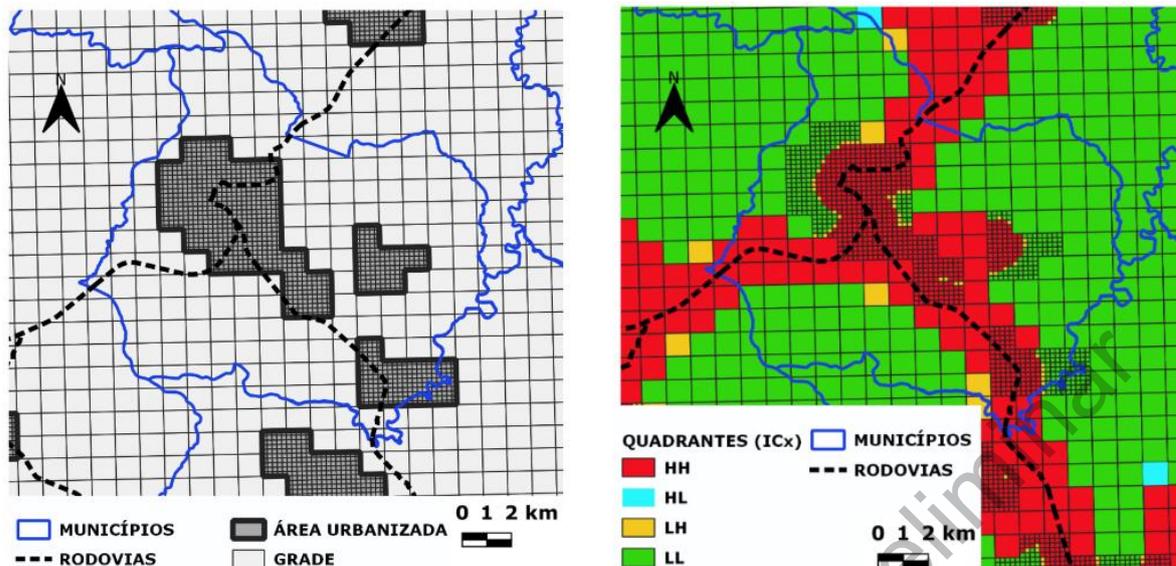


Figura 2: À esquerda, estrutura espacial da grade para a obtenção da variável %AU para um dado município; à direita, detalhe da grade com as células classificadas segundo os quadrantes do gráfico de espalhamento de Moran para o indicador de oferta de infraestrutura rodoviária

A Figura 3 apresenta a estrutura da árvore de decisão do primeiro modelo (%HH e %AU). Observa-se que %HH é a variável da primeira subdivisão, separando o primeiro grupo de municípios que seriam definidos como pertencentes a uma RUH caso a sua %HH fosse maior do que 55,87%. Na sequência, caso a %HH fosse menor do que 55,87%, o município poderia ainda ser definido como pertencente a uma RUH caso a sua %AU fosse maior do que 30,35%. Caso contrário, o município não pertenceria a uma RUH.

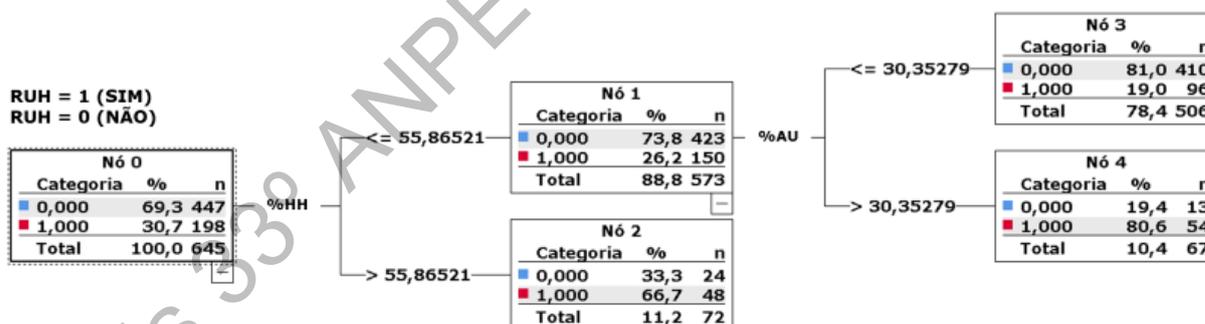


Figura 3: Árvore de decisão do primeiro modelo, constituído das variáveis %HH e %AU

O mapa da Figura 4 ilustra a aplicação das regras obtidas pelo primeiro modelo, destacando-se os municípios que pertenceriam a uma RUH. As atuais definições das RUHs oficiais também estão sobrepostas ao mapa, podendo-se verificar a coincidência entre a situação oficial e a situação definida pelo primeiro modelo. A Tabela 1 apresenta tais coincidências em termos quantitativos, com as porcentagens de: municípios que pertencem a uma RUH oficial e também atendem às regras do primeiro modelo, municípios que pertencem a uma RUH mas não atendem às regras do referido modelo, municípios que não pertencem a uma RUH e também não atendem às regras, e municípios que não pertencem a uma RUH mas atendem às regras.

De maneira semelhante para o segundo modelo, que foi construído com base nas variáveis %LL e %AU, verifica-se pela Figura 5 que %LL é a variável da primeira subdivisão, separando o

primeiro grupo de municípios que seriam definidos como pertencentes a uma RUH caso a sua %LL fosse menor do que 36,55%. Na sequência, caso a %LL fosse maior do que 36,55%, o município poderia ainda ser definido como pertencente a uma RUH caso a sua %AU fosse maior do que 30,39%. Caso contrário, o município não pertenceria uma RUH. Na Figura 6 está o mapa que ilustra a aplicação das regras obtidas para esse modelo. Neste caso também é possível identificar visualmente a coincidência entre a situação oficial e a situação definida pelo segundo modelo, complementada pelos resultados quantitativos apresentados na Tabela 2.

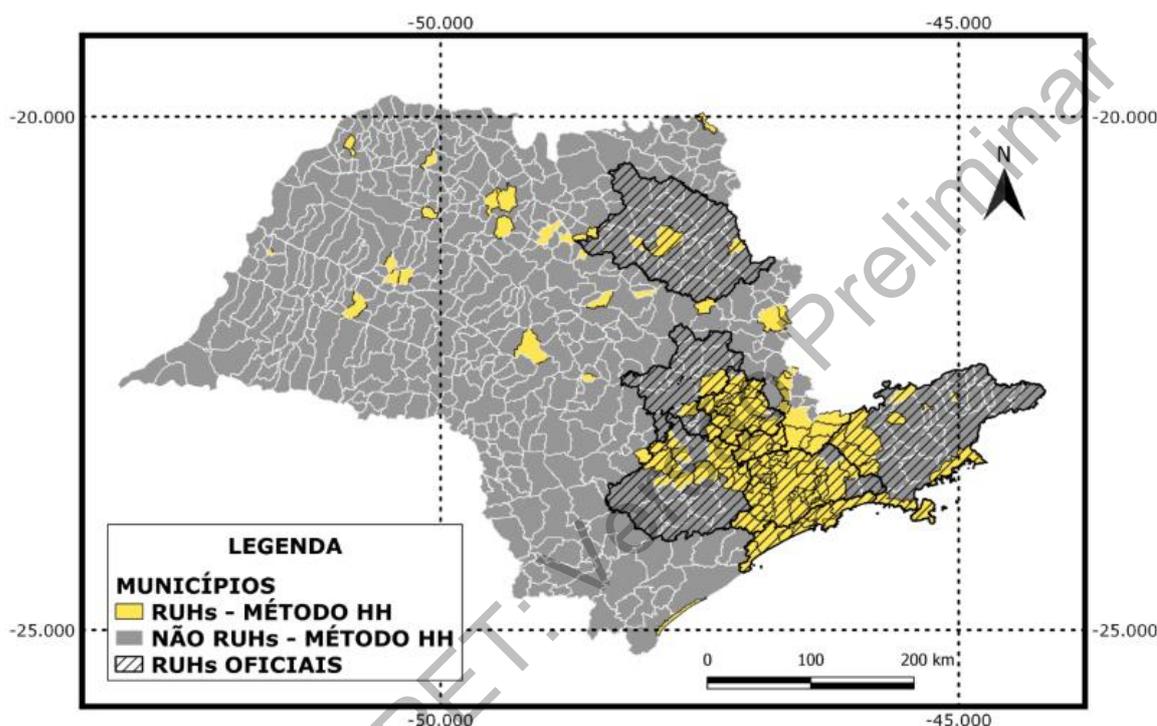


Figura 4: Mapa temático contendo as RUHs previstas pelo primeiro modelo e o contorno das RUHs oficiais atuais

Tabela 1: Resultados quantitativos do primeiro modelo

| | RUH = sim (modelo) | RUH = não (modelo) |
|---|------------------------------|------------------------------|
| RUH = sim (oficial); n = 198 | 51,5% (102/198) ¹ | 48,5% (96/198) ² |
| RUH = não (oficial); n = 447 | 8,3% (37/447) ² | 91,7% (410/447) ¹ |
| Porcentagem correta global ³ | 79,4% (512/645) | |

¹ Porcentagens de acertos do modelo; ² Porcentagens de erros do modelo; ³ Obtida por meio da razão entre as estimativas corretas (102+410) e a quantidade total de municípios (645).

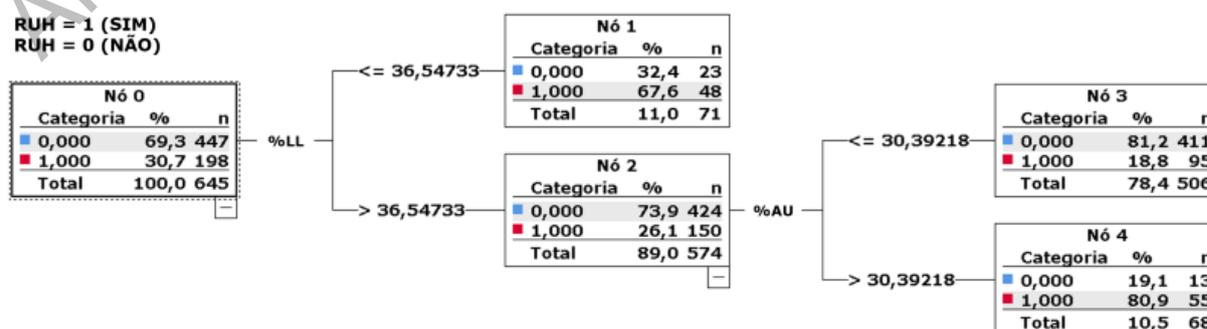


Figura 5: Árvore de decisão do segundo modelo, constituído das variáveis %LL e %AU

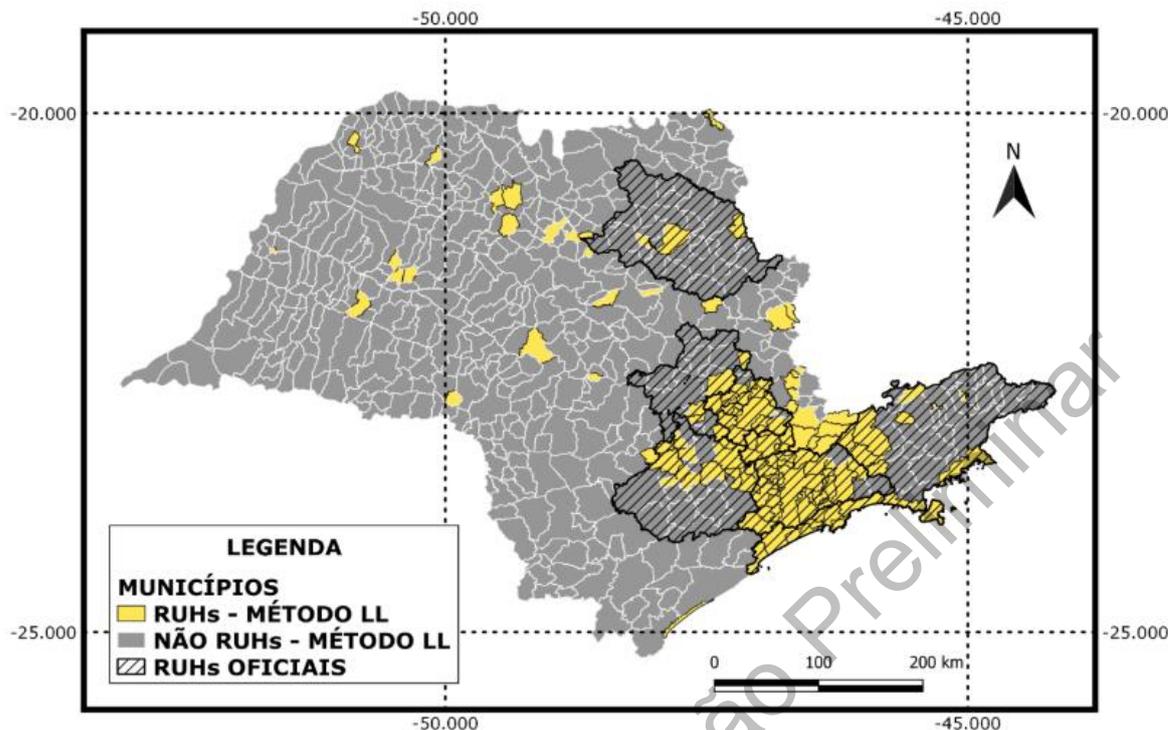


Figura 6: Mapa temático contendo as RUHs previstas pelo segundo modelo e o contorno das RUHs oficiais atuais

Tabela 2: Resultados quantitativos do segundo modelo

| | RUH = sim (modelo) | RUH = não (modelo) |
|---|------------------------------|------------------------------|
| RUH = sim (oficial); n = 198 | 52,0% (103/198) ¹ | 48,0% (95/198) ² |
| RUH = não (oficial); n = 447 | 8,1% (36/447) ² | 91,9% (411/447) ¹ |
| Porcentagem correta global ³ | 79,7% (514/645) | |

¹ Percentagens de acertos do modelo; ² Percentagens de erros do modelo; ³ Obtida por meio da razão entre as estimativas corretas (103+411) e a quantidade total de municípios (645).

4. CONCLUSÃO

O objetivo deste trabalho foi apresentar um método para a definição de RUHs com base em árvores de decisão, combinando-se dados de oferta de infraestrutura rodoviária com a grade estatística do IBGE. Como parte desse objetivo está a minimização do problema da falácia ecológica, proporcionada pela análise inicial dos dados em uma escala mais desagregada (neste caso, a da grade). Em relação às variáveis consideradas, a porcentagem de área urbanizada em relação à área total do município (%AU) se mostrou bastante relevante. Sua inclusão foi fundamental para o bom desempenho dos modelos, mostrando-se, ao mesmo tempo, consistente com o próprio processo de urbanização e formação de RUHs.

Quanto às variáveis relacionadas ao indicador de oferta de infraestrutura rodoviária, obtidas por meio da técnica ESDA (%HH, %LL, %LH e %HL), apenas as variáveis %HH e %LL mostraram-se significativas para os modelos, e ainda de maneira separada entre elas. Esse fato pareceu ser coerente, pois se %HH é muito grande em um município, a %LL tende a ser pequena, e vice-versa. Assim, a obtenção de dois modelos separados, sendo o primeiro baseado apenas em %HH e em %AU, e o segundo baseado apenas em %LL e em %AU pareceu razoável. Além disso, como os resultados dos dois modelos são muito próximos, pode-se dizer que eles são

equivalentes para a questão da definição de RUHs. Cabe destacar ainda a relação positiva da inclusão de dados de oferta de infraestrutura rodoviária na referida questão.

Em relação ao emprego das árvores de decisão, essa técnica mostrou-se satisfatória ao fornecer as regras de decisão para a definição de RUHs conforme proposto neste trabalho. Cabe ressaltar que não se esgotam aqui as análises de outras técnicas e de outras variáveis. Por exemplo, redes neurais artificiais, redes de classificação com inferência Bayesiana, regressão logística etc. poderiam ser exploradas. Além disso, a inclusão da distribuição populacional seria outra variável evidente, proporcionando um próximo passo na metodologia deste trabalho. Tal variável não foi considerada aqui somente pelo fato de que havia o interesse em analisar primeiramente a oferta de infraestrutura rodoviária de maneira independente da densidade populacional. Entretanto, a análise conjunta dessas variáveis deverá ser reportada em oportunidades futuras.

Em resumo, a proposta deste estudo visou contribuir também com uma discussão apresentada por Moura e Carvalho (2012) no sentido de que diversas RUHs existentes no Brasil foram instituídas sem uma dinâmica metropolitana expressiva. Conforme a Constituição Federal de 1988, cabe aos estados elaborarem seus próprios critérios para a definição e instalação de RUHs, decorrendo em grandes divergências e sugerindo definições baseadas em relações político-administrativas predominantemente. Sendo assim, se por um lado a instalação de RUHs torna mais favorável a criação e aplicação de políticas de planejamento e gestão urbanas integradas às questões econômicas, sociais e político-administrativas dos municípios que compõem essas RUHs, por outro lado existe um grande desafio para planejadores e gestores urbanos quanto à questão dos critérios para a definição das RUHs.

Agradecimentos

Os autores agradecem às agências CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, processos 303276/2018-5 e 400617/2016-1) e FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo, processo 2018/19281-3) pelo apoio concedido em diferentes fases da pesquisa que deu origem a este trabalho.

REFERÊNCIAS BIBLIOGRÁFICAS

- Aguiar, L. L. e G. G. Manzato (2017) Análise exploratória de dados de viagens pendulares visando a definição de Regiões Urbanas Homogêneas. *Transportes*, v. 25, n. 3, p. 1-11. DOI:10.14295/transportes.v25i3.1305
- Aguiar, L. L.; G. G. Manzato e A. N. Rodrigues da Silva (2017) Patterns of commuting flows for delimitating Functional Urban Regions in the state of São Paulo, Brazil. *Proceedings of the 15th International Conference on Computers in Urban Planning and Urban Management (CUPUM)*, Adelaide, Australia.
- Ajauskas, R.; G. G. Manzato e A. N. Rodrigues da Silva (2012) The definition of Functional Urban Regions: Validation of a set of spatial models with recent census data and analysis of an additional model specification. *Proceedings of CAMUSS, the International Symposium on Cellular Automata Modeling for Urban Spatial Systems*, Porto, Portugal, p. 91-104.
- Anselin, L. (1995) Local Indicators of Spatial Association – LISA. *Geographical Analysis*, v. 27, p. 93-115. DOI:10.1111/j.1538-4632.1995.tb00338.x.
- Anselin, L. (1998) Exploratory spatial data analysis in a geocomputational environment. In: P. A. Longley; S. M. Brooks; R. McDonnell e W. Macmillian (eds.) *Geocomputation: A Primer*. NY: Wiley and Sons, p. 77-94.
- Arellano Ríos, A. (2014a) La definición jurídica del fenómeno metropolitano en el ámbito subnacional mexicano. *Opinión Jurídica*, v. 13, n. 26, p. 91-108.
- Arellano Ríos, A. (2014b) La coordinación metropolitana en el ámbito subnacional mexicano: un análisis institucional. *Documentos y Aportes en Administración Pública y Gestión Estatal (DAAPGE)*, v. 14, n. 23, p. 33-70.
- Arellano Ríos, A. (2015) Metropolitan coordination in Mexico. *Current Urban Studies*, v. 3, p. 11-17. DOI:10.4236/cus.2015.31002.
- Bereitschaft, B. e K. Debbage (2014) Regional variations in urban fragmentation among U.S. metropolitan and megapolitan areas. *Applied Spatial Analysis and Policy*, v. 7, p. 119-147. DOI:10.1007/s12061-013-9092-9.
- Boarnet, M. G. e A. F. Haughwout (2000) Do highways matter? Evidence and policy implications of highways' influence on metropolitan development. The Brookings Institution Center on Urban and Metropolitan Policy

- (Discussion Paper) <<http://www.brookings.edu/research/reports/2000/08/highways-boarnet>>
- Bode, E. (2008) Delineating metropolitan areas using land prices. *Journal of Regional Science*, v. 48, n. 1, p. 131-163. DOI:10.1111/j.1467-9787.2008.00544.x.
- Cheshire, P. C. e D. G. Hay (1989) *Urban problems in Western Europe: an economic analysis*. Unwin Hyman, London.
- Coombes, P. P. e H. G. Overman (2004) The spatial distribution of economic activities in the European Union. In: J. V. Henderson e J. F. Thisse (eds.) *Handbook of Urban and Regional Economics: Cities and Geography*. North Holland, Amsterdam, v. 4, p. 2845-2909.
- Gouvêa, R. G. (2005) *A questão metropolitana no Brasil*. FGV, Rio de Janeiro.
- Guzik, R.; A. Kołós e K. Gwosdz (2017) Interconnections in public transport as a method for delimiting urban functional areas and the settlement hierarchy in Poland. *Regional Statistics*, v. 7, n. 1, p. 063-077. DOI:10.15196/RS07104
- Hornák, M. e S. Kraft (2015) Functional Transport Regions in Slovakia defined by passenger-car traffic flows. *Mitteilungen der Österreichischen Geographischen Gesellschaft*, v. 157, n. 1, p. 109-128. DOI:10.1553/moegg157s109
- Huang, X.; Q. Lu e L. Zhang (2014) A multi-index learning approach for classification of high-resolution remotely sensed images over urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 90, p. 36-48. DOI:10.1016/j.isprsjprs.2014.01.008.
- Jargowsky, P. A. (2005) Ecological fallacy. *Encyclopedia of Social Measurement*, v. 1, p. 715-722.
- Kamiński, B.; M. Jakubczyk e P. Szufel (2017) A framework for sensitivity analysis of decision trees. *Central European Journal of Operations Research*. DOI:10.1007/s10100-017-0479-6.
- Kauffmann, A. (2012) *Delineation of city regions based on commuting interrelations: The example of large cities in Germany*. IWH-Diskussionspapiere, n. 2012.4. Available at: <http://nbn-resolving.de/urn:nbn:de:101:1-201212176768>.
- Klapka, P. e M. Halás (2016) Conceptualising patterns of spatial flows: Five decades of advances in the definition and use of functional regions. *Moravian Geographical Reports*, v. 24, n. 2, p. 2-11. DOI:10.1515/mgr-2016-0006
- Konjar, M.; A. Lisec e S. Drobne (2010) Methods for delineation of functional regions using data on commuters. *Proceedings of the 13th AGILE International Conference on Geographic Information Science*. Guimarães, Portugal.
- Kourtít, K.; P. Nijkamp e M. D. Partridge (2015) Challenges of the new urban world. *Applied Spatial Analysis and Policy*, v. 8, n. 3, p. 199-215. DOI:10.1007/s12061-015-9155-1
- Kraft, S. e M. Marada (2017) Delimitation of functional transport regions: understanding the transport flows patterns at the micro-regional level. *Geografiska Annaler: Series B, Human Geography*. DOI:10.1080/04353684.2017.1291741.
- Magalhães, M. T. (2016) Spatial coverage index for assessing national and regional transportation infrastructures. *Journal of Transport Geography*, v. 56, p. 53-61. DOI: 10.1016/j.jtrangeo.2016.08.015.
- Manley, E. (2014) Identifying functional urban regions within traffic flow. *Regional Studies, Regional Science*, v. 1, n. 1, p. 40-42. DOI:10.1080/21681376.2014.891649
- Manzato, G. G. e A. N. Rodrigues da Silva (2010) Spatial-temporal combination of variables for monitoring changes in metropolitan areas. *Applied Spatial Analysis and Policy*, v. 3, n. 1, p. 25-44. DOI:10.1007/s12061-009-9028-6.
- Manzato, G. G.; I. Baria e A. N. Rodrigues da Silva (2007) A GIS-based comparison of methodologies for the definition of metropolitan areas in a developing country. *Proceedings of the 10th International Conference on Computers in Urban Planning and Urban Management*, Iguazu Falls, Brazil.
- Manzato, G. G.; R. S. Dias e A. N. Rodrigues da Silva (2015) Capacity of roadway infrastructure and its relation with Functional Urban Regions. *Proceedings of the 14th International Conference on Computers in Urban Planning and Urban Management*, Cambridge MA, USA.
- Matiolli, J. A. C.; M. A. de Oliveira Junior e G. G. Manzato (2017) Modelagem espacial para a definição de regiões urbanas homogêneas incorporando recentes dados demográficos e de oferta de infraestrutura rodoviária. *Transportes*, v. 25, n. 3, p. 12-23. DOI:10.14295/transportes.v25i3.1306.
- Moisés, O. A.; V. Royuela e V. Xavier (2017) Computing functional urban areas using a hierarchical travel time approach: An applied case to Ecuador. University of Barcelona, Research Institute of Applied Economics.
- Moura, R. e I. Carvalho (2012) *Estatuto da Metrópole: onde está a região metropolitana?*. Observatório das Metrópoles, Available at: http://observatoriodasmetrolopes.net/index.php?option=com_k2&view=item&id=455%3Aestatuto-da-metr%C3%B3pole-onde-est%C3%A1-a-regi%C3%A3o-metropolitana%3F&Itemid=165&lang=pt. Accessed April 15, 2014.
- Niemeyer, J.; F. Rottensteiner e U. Soergel (2014) Contextual classification of Lidar data and building object detection in urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 87, p. 152-165.

- DOI:10.1016/j.isprsjprs.2013.11.001.
- OECD (2016) *Regions at a Glance*, OECD Publishing, Paris. http://dx.doi.org/10.1787/reg_glance-2016-en
- Office of Management and Budget (1998) Alternative approaches to defining metropolitan and non-metropolitan areas, *Federal Register*, v. 63, n. 244, December 21, 1998.
- Office of Management and Budget (2010) 2010 Standards for delineating metropolitan and micropolitan statistical areas; Notice. *Federal Register*, v. 75, n. 123, June 28, 2010.
- Quinlan, J. R. (1987) Simplifying decision trees. *International Journal of Man-Machine Studies*. v. 27, n. 3, p. 221. DOI:10.1016/S0020-7373(87)80053-6.
- Pereira, H. T. S. e A. N. Rodrigues da Silva (2010) Comparing spatial analysis methods for the definition of Functional Urban Regions - The case of Bahia, Brazil. *Proceedings of the 10th International Conference on Design and Decision Support Systems in Architecture and Urban Planning*, Eindhoven, The Netherlands.
- Ramos, R. A. R. e A. N. Rodrigues da Silva (2003) A data-driven approach for the definition of metropolitan regions. *Proceedings of the 8th International Conference on Computers in Urban Planning and Urban Management*, Sendai, Japan.
- Ramos, R. A. R. e A. N. Rodrigues da Silva (2007) A spatial analysis approach for the definition of metropolitan regions - the case of Portugal. *Environment and Planning B: Planning and Design*, v. 34, n. 1, p. 171-185. DOI:10.1068/b31117.
- Ramos, R. A. R.; A. N. Rodrigues da Silva e V. P. Miranda (2004) A comparison of two methods for the definition of regional metropolitan areas through an application in the north of Portugal. *Proceedings of the 44th European Congress of the European Regional Science Association*, Porto, Portugal.
- Rodrigues da Silva, A. N.; G. G. Manzato e H. T. S. Pereira (2014) Defining Functional Urban Regions in Bahia, Brazil, using roadway coverage and population density variables. *Journal of Transport Geography*, v. 36, p. 79-88. DOI:10.1016/j.jtrangeo.2014.03.001.
- Sahoo, S. N. e S. Pekkat (2014) Determination of urbanization based on imperviousness. *Urban Design and Planning*, v. 167, n. DP2, p. 49-57. DOI:10.1680/udap.13.00027.
- Santos, S. P.; G. G. Manzato e A. N. Rodrigues da Silva (2014) Estratégias para a determinação da densidade populacional visando a definição de regiões urbanas homogêneas por meio de técnicas de análise espacial. *Anais do 6º Congresso Luso Brasileiro para o Planejamento Urbano, Regional, Integrado e Sustentável*, Lisboa, Portugal.
- São Paulo (1989) Constituição do Estado de São Paulo. Governo do Estado de São Paulo. Disponível em: www.legislacao.sp.gov.br/legislacao/index.htm.
- Soares, E.; R. Figueiredo e F. Vala (2017) Defining Labour Market Areas and its relevance from a statistical perspective: The Portuguese case. *Statistical Journal of the IAOS*, v. 33, n. 3, p. 615-625. DOI:10.3233/SJI-170381.
- Souza, V. L. C. e G. G. Manzato (2017) Comparação de técnicas de análise espacial aplicadas a setores censitários visando a identificação de Regiões Urbanas Homogêneas. *Anais do SINGEORB 2017 - I Simpósio Nacional de Gestão e Engenharia Urbana*, São Carlos.
- UN (2007) United Nations, Population fund, state of world population 2007, New York: UN.
- Weber, C. (2001) Urban agglomeration delimitation using remote sensing data. In: J. P. Donnay, M. J. Barnsley e P. A. Longley (eds.) *Remote sensing and urban analysis*. Taylor & Francis, London, p. 145-159.
- Williams, A. M.; J. Foord e J. Mooney (2012) Human mobility in functional urban regions: understanding the diversity of mobilities. *International Review of Sociology*, v. 22, n. 2, p. 191-209. DOI:10.1080/03906701.2012.696961.

Caio Cesar dos Anjos Carneiro (caio.cacarneiro@gmail.com)
Gustavo Garcia Manzato (gustavo.manzato@unesp.br)
Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP), Faculdade de Engenharia de Bauru
Av. Eng. Luiz Edmundo C. Coube, 14-01 - 17033-360 Bauru, SP, Brasil