

# Modelagem espacial para a definição de regiões urbanas homogêneas incorporando recentes dados demográficos e de oferta de infraestrutura rodoviária

João Antonio Camargo Matioli<sup>1</sup>, Mauro Antonio de Oliveira Junior<sup>2</sup>, Gustavo Garcia Manzato<sup>3</sup>✉

<sup>1</sup>Departamento de Engenharia Civil e Ambiental, UNESP, joao\_antonio79@hotmail.com

<sup>2</sup>Departamento de Engenharia Civil e Ambiental, UNESP, mauroaojunior@yahoo.com.br

<sup>3</sup>Departamento de Engenharia Civil e Ambiental, UNESP, gusmanzato@gmail.com

## Recebido:

17 de janeiro de 2017

## Aceito para publicação:

27 de setembro de 2017

## Publicado:

29 de outubro de 2017

## Editor de área:

Cira Souza Pitombo

## Palavras-chaves:

Regiões urbanas homogêneas,  
Oferta de infraestrutura rodoviária,  
Modelagem espacial.

## Keywords:

Functional urban regions,  
Road infrastructure supply,  
Spatial modeling.

DOI:10.14295/transportes.v25i3.1306

## RESUMO

O objetivo deste trabalho foi desenvolver os modelos espaciais para a definição de Regiões Urbanas Homogêneas (RUHs) propostos por Manzato e Rodrigues da Silva (2010) e Ajauskas *et al.* (2012) por meio da incorporação de dados demográficos e de oferta de infraestrutura rodoviária referentes ao ano de 2010, disponibilizados recentemente. Isso permitiu uma nova estimativa dos parâmetros com dados adicionais, promovendo uma atualização desses modelos. A metodologia empregada se baseou em recursos de estatística espacial e modelagem espacial. Os resultados dos modelos aqui desenvolvidos demonstraram que a incorporação dos novos dados garantiu uma consistência equivalente aos modelos anteriormente propostos pelos referidos autores. Em particular, o desempenho de dois modelos, ambos incorporando a oferta de infraestrutura rodoviária, foi superior nas análises envolvendo as classificações de municípios no quadrante HH do gráfico de Moran. Isso demonstra o potencial da abordagem baseada na combinação de indicadores demográficos e de oferta de infraestrutura de transportes para a definição de RUHs.

## ABSTRACT

The objective of this study was to develop spatial models for the definition of Functional Urban Regions (FURs) proposed by Manzato and Rodrigues da Silva (2010) and Ajauskas *et al.* (2012) through the incorporation of demographic and road infrastructure supply data related to the year 2010, which became available recently. This enabled a new estimation of the parameters with additional data, providing an update of those models. The methodology was based on spatial statistics and spatial modeling techniques. The results of the models developed here demonstrated that the incorporation of the new data secured a consistency equivalent to the previous models proposed by the cited authors. In particular, the performance of two models, both incorporating road infrastructure supply, was higher in the analyses involving the municipalities classified in the quadrant HH of the Moran scatterplot. This demonstrates the potential of the approach based on the combination of demographic and transport infrastructure supply indicators for the definition of FURs.



## 1. INTRODUÇÃO

De acordo com recentes estudos das Nações Unidas (UN, 2015), 54 % da população mundial atualmente vive em áreas urbanas. Ou seja, há mais pessoas nessas áreas do que em áreas rurais. Esses mesmos estudos mostram uma projeção da população mundial urbana, que deve chegar a 66 % até 2050. Consequentemente, profundas mudanças tanto no tamanho quanto na distribuição espacial da população global são esperadas.

Historicamente, a busca por melhores condições de vida é a principal motivação da migração das pessoas para as cidades (ver, por exemplo Rodriguez-pose e Ketterer, 2012). De fato, as cidades concen-

tram a maioria das atividades econômicas e possuem as melhores redes de infraestrutura (transportes, telecomunicações, energia elétrica, saúde, educação, lazer, etc.). Isso oferece grandes oportunidades para o desenvolvimento urbano, mas ao mesmo tempo provoca uma enorme pressão nessas áreas, pois as cidades são centros de atividades humanas com interações sociais e ambientais bastante intensas (Kourtit e Nijkamp, 2013). Conforme Rodrigues da Silva *et al.* (2008), essas interações são uma consequência das relações espaciais, econômicas, políticas, sociais e demográficas existentes nessas áreas urbanizadas, podendo-se observar vários esforços em elaborar modelos capazes de representar as inter-relações presentes nessas estruturas multidimensionais e complexas.

Uma questão crucial presente nas investigações acerca desse tema reside no fato de que a ocupação do território geralmente não se restringe a apenas um município e, muitas vezes, o processo de urbanização acaba abrangendo mais de um município. Como as referidas interações normalmente dependem da estrutura hierárquica das cidades e das características das infraestruturas disponíveis, regiões com maior poder de atração de pessoas e de bens podem se destacar sobre outras em um cenário regional ou nacional. Um resultado desse processo é a formação de Regiões Urbanas Homogêneas (RUHs), denominação dada neste contexto para contemplar, de maneira geral, as regiões metropolitanas, aglomerações urbanas e conurbações. Para fins de políticas de desenvolvimento e resolução de problemas na escala metropolitana, planejadores e gestores urbanos encontram o desafio a respeito da delimitação dessas RUHs. Por um lado, os limites administrativos dessas áreas geralmente estão além dos limites individuais dos municípios que as compõem. Mas por outro, eles também não correspondem ou não são suficientes para atender à subdivisão administrativa superior, como os estados, por exemplo.

Breitung (2011) sugere que a análise da delimitação das RUHs deveria ser feita de acordo com cinco diferentes, embora inter-relacionados, aspectos: questões políticas e administrativas; características físicas ou morfológicas; questões socioespaciais e culturais; perspectivas psicológicas e comportamentais; e relações funcionais e de redes de cooperação. A literatura internacional apresenta diversos estudos contemplando parte desses aspectos. Exemplos 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; Debbage *et al.*, 2017); e análise de movimentos pendulares (Cheshire e Hay, 1989; *Office of Management and Budget*, 2000; Kauffmann, 2012; Williams *et al.*, 2012; Aguiar e Manzato, 2017; Aguiar *et al.*, 2017).

Por outro lado, conforme Kourtit *et al.* (2015), esses aspectos são muitas vezes difíceis de serem mensurados ou podem apresentar definições ambíguas ou imprecisas. Nesse caso, medidas de distribuição populacional (por exemplo, a densidade populacional) ainda são os indicadores mais coerentes para se medir o grau de urbanização (UN, 2007). Isso também foi sugerido pelo *Office of Management and Budget* (1998). Tradicionalmente, o censo populacional é realizado de maneira regular em praticamente todos os países e seus resultados apresentam um elevado grau de confiabilidade. Assim, a densidade populacional serviria como uma alternativa para a definição de RUHs na falta de outras medidas consistentes e confiáveis nacionalmente. Isso foi mostrado por Ramos e Rodrigues da Silva (2003 e 2007), Ramos *et al.* (2004), Manzato *et al.* (2007) e Santos *et al.* (2014), que exploraram a referida variável com técnicas de análise espacial, como a estatística espacial e a modelagem espacial.

Seguindo essa linha de pesquisa com a aplicação das técnicas de análise espacial, alternativas de incorporação de dados oferta de infraestrutura de transportes emergiram visando medir a interação entre regiões. Isto é, se a formação e o desenvolvimento de RUHs pressupõem interações entre regiões, essas interações dependem de infraestruturas. No caso aqui analisado, essas infraestruturas correspondem sobretudo à de transportes. Assim, quanto mais intensas essas interações ocorrem, maior será a demanda por infraestrutura. Conseqüentemente, assume-se que a oferta de infraestrutura acompanhe essa demanda e que o nível de oferta em uma determinada região possa ser utilizado para a definição de RUHs. Isso foi investigado por Manzato *et al.* (2006) e Dias *et al.* (2014) com a aplicação de técnicas

de estatística espacial sobre indicadores de oferta de infraestrutura rodoviária. Diante dos resultados obtidos, os autores sugeriram que um indicador para tal finalidade deveria ser uma composição de indicadores de oferta de infraestrutura de transportes e populacionais. Isso foi baseado nas evidências sobre o papel da oferta de infraestrutura de transportes no desenvolvimento urbano e regional (ver, por exemplo, os trabalhos de Lin, 1999; Boarnet e Haughwout, 2000; Baum-Snow, 2010). Essa hipótese foi explorada em estudos subsequentes (Manzato e Rodrigues da Silva, 2006, 2007 e 2010; Pereira e Rodrigues da Silva, 2010; Ajauskas *et al.*, 2012 e Rodrigues da Silva *et al.*, 2014), cujos resultados demonstraram que tal composição apresenta um forte potencial para a definição de RUHs.

Baseando-se nesses estudos mencionados, o objetivo deste trabalho foi desenvolver os modelos espaciais para a definição de RUHs propostos particularmente por Manzato e Rodrigues da Silva (2010) e Ajauskas *et al.* (2012) por meio da incorporação de dados demográficos e de oferta de infraestrutura rodoviária referentes ao ano de 2010. Esses dados foram disponibilizados recentemente por meio do Instituto Brasileiro de Geografia e Estatística (IBGE, 2010) e do Departamento de Estradas de Rodagem do Estado de São Paulo (DER-SP, 2010). Isso permitiu uma nova estimativa dos parâmetros dos modelos espaciais com dados adicionais, promovendo uma atualização desses modelos. A metodologia empregada se baseou em recursos de estatística espacial, como a técnica ESDA (do inglês, *Exploratory Spatial Data Analysis*), e de modelagem espacial utilizando princípios de *Cellular Automata* (CA). O artigo está estruturado da seguinte maneira: na seção 2 é apresentada a metodologia juntamente com a descrição do estudo de caso desenvolvido para o estado de São Paulo. Na sequência, os resultados são discutidos na seção 3. Por fim, na seção 4 apresentam-se as conclusões, seguidas das referências bibliográficas citadas neste trabalho.

## 2. METODOLOGIA

O presente artigo segue a metodologia proposta por Manzato e Rodrigues da Silva (2010) e Ajauskas *et al.* (2012) que compreende, mais especificamente, modelos espaciais elaborados para a definição de RUHs. Esses modelos utilizam a técnica ESDA e princípios de CA (ver, por exemplo, White e Engelen, 1993a e 1993b; Batty e Xie, 1994; Cecchini, 1996; Batty *et al.*, 1997; Clarke *et al.*, 1997) aplicados em uma composição de indicadores populacionais e de oferta de infraestrutura de transportes. O indicador populacional utilizado foi a densidade populacional calculada para cada município do estado de São Paulo. O indicador de oferta de infraestrutura de transportes se referiu ao índice de cobertura espacial de estrutura de circulação viária (conforme Magalhães *et al.*, 2004), mostrado na Equação 1. Este índice diz respeito à área influenciada por um sistema de transporte (no caso deste estudo, o rodoviário) e é composto por 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.

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

em que:  $IC_x$ : índice de cobertura espacial de estrutura de circulação viária para zona  $x$ ;  
 $\gamma(i)$ : função que determina o peso atribuído à  $i$ -ésima faixa de cobertura (*buffer*), tal que  $\gamma(i) \in [0,1]$ , tendo sido aqui empregado uma função polinomial de 5º grau;  
 $A_i$ : área da  $i$ -ésima faixa contida em uma zona  $x$ ;  
 $A_x$ : área da zona  $x$  (no caso deste estudo, municípios do estado de São Paulo);  
 $n$ : número de faixas (foram adotadas neste estudo 10 faixas com 4 km de equidistância cada).

No caso da aplicação da técnica ESDA, cada zona (ou município, no caso deste estudo) é classificada por meio de dois parâmetros. O primeiro,  $Z$ , é função da diferença entre o valor de um determinado

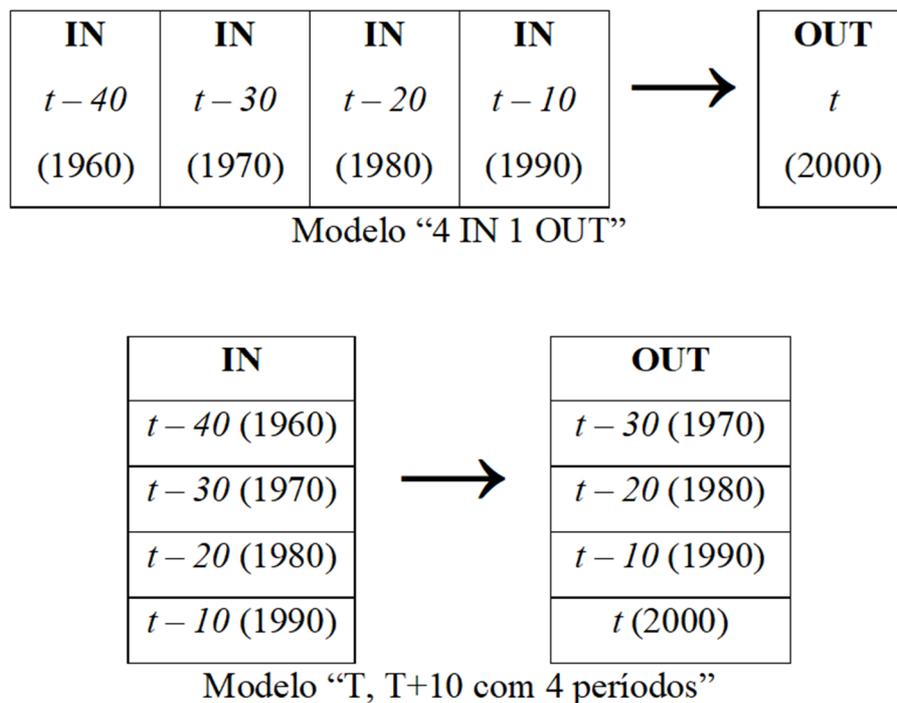
atributo (por exemplo, a densidade populacional) e a média global desse atributo. O segundo,  $Wz$ , é função da diferença da média do atributo nas zonas adjacentes e a média global desse atributo. Os resultados dessa classificação, que podem ser representados em quatro quadrantes do gráfico de espalhamento de Moran e também em mapas (chamados de “*Box Maps*”), podem ser analisados conforme se descreve a seguir.

No quadrante Q1 (ou HH, de *High-High*) estão representadas as zonas em que o atributo da zona possui valor superior à média global e que o valor médio das áreas adjacentes também é superior à média global. Para o caso da densidade populacional, centros de regiões metropolitanas são geralmente classificadas nesse quadrante. No quadrante Q2 (ou LL, de *Low-Low*) estão representadas as zonas em que o atributo da zona possui valor inferior à média global e que o valor médio das áreas adjacentes também é inferior à média global. Para o caso da densidade populacional, a maior parte dos municípios é normalmente classificada nesse quadrante e se refere a cidades pequenas e distantes dos grandes centros urbanos. No quadrante Q3 (ou LH, de *Low-High*) estão representadas as zonas em que o atributo da zona possui valor inferior à média global e que o valor médio das áreas adjacentes é superior à média global. Para o caso da densidade populacional, os municípios situados nas periferias de regiões metropolitanas e de grandes cidades são classificadas nesse quadrante. Finalmente, no quadrante Q4 (ou HL, de *High-Low*) estão representadas as zonas em que o atributo da zona possui valor superior à média global e que o valor médio das áreas adjacentes é inferior à média global. Para o caso da densidade populacional, as grandes cidades do território geralmente isoladas e afastadas das regiões metropolitanas são comumente classificadas nesse quadrante, representando polos regionais.

Cabe ressaltar que os quadrantes do gráfico de Moran são obtidos tanto para a densidade populacional ( $PDx$ ) como para o índice de oferta rodoviária ( $ICx$ ), dando origem às variáveis  $QPDx$  e  $QICx$ , respectivamente. Além dessas quatro variáveis, na modelagem espacial baseada em CA proposta por Manzato e Rodrigues da Silva (2010) e Ajauskas *et al.* (2012), as seguintes variáveis adicionais também foram consideradas: média da densidade populacional calculada para os municípios vizinhos de cada município  $x$  ( $PDj$ ); quantidade de municípios vizinhos classificados em cada quadrante do gráfico de Moran obtido para  $PDx$ , isto é,  $n_{HH_x^{PDx}}$ ,  $n_{LL_x^{PDx}}$ ,  $n_{LH_x^{PDx}}$  e  $n_{HL_x^{PDx}}$ . Em resumo, nove variáveis foram definidas para tais modelos espaciais. Os dados para a determinação dessas variáveis foram obtidos junto ao IBGE e ao DER-SP, tendo como período de análise os anos de 1960, 1970, 1980, 1991 (aqui considerado como 1990), 2000 e 2010. Em particular, a inclusão dos dados de 2010 disponibilizados recentemente proporcionou o diferencial deste estudo comparado aos desenvolvidos por Manzato e Rodrigues da Silva (2010) e Ajauskas *et al.* (2012), embora a estrutura básica dos modelos tenha permanecido a mesma.

Dois estruturas de modelos foram inicialmente desenvolvidas. A primeira, chamada “3 IN 1 OUT”, tem como dados de entrada três períodos subsequentes e igualmente espaçados. O dado de saída se refere ao período seguinte. Por exemplo, caso o dado de saída se refira ao período “ $t$ ” e os intervalos entre os dados sejam de 10 anos, os dados de entrada devem se referir aos períodos “ $t - 30$ ”, “ $t - 20$ ” e “ $t - 10$ ” anos. Essa estrutura foi elaborada por Manzato e Rodrigues da Silva (2010) em função da disponibilidade de dados na época de desenvolvimento do estudo, que compreendia os anos de 1960, 1970, 1980, 1990 e 2000. Como o processo de modelagem é geralmente composto por uma fase de estimativa de parâmetros e uma fase de validação desses parâmetros, os autores utilizaram na primeira fase os anos de 1960, 1970, 1980 e 1990 (sendo 1960, 1970 e 1980 como dados de entrada e 1990 como dado de saída). Na segunda fase (validação), entraram os anos de 1970, 1980, 1990 e 2000, sendo 1970, 1980 e 1990 como dados de entrada e 2000 como dado de saída. Sendo assim, com a disponibilização dos dados de 2010, essa primeira estrutura denominada “3 IN 1 OUT” passou a ser chamada de “4 IN 1 OUT”. De maneira análoga, se o dado de saída corresponde ao período “ $t$ ”, os dados de entrada devem se referir aos períodos “ $t - 40$ ”, “ $t - 30$ ”, “ $t - 20$ ” e “ $t - 10$ ” anos. Portanto, para a fase de estimativa de parâmetros foram utilizados os anos de 1960, 1970, 1980 e 1990 como dados de entrada e o ano de 2000 como dado de saída. Para a fase de validação, os dados de entrada se referiram aos anos de 1970, 1980, 1990 e 2000, e como dado de saída foi utilizado o ano de 2010.

A segunda estrutura de modelos é chamada de “T, T + 10”. Aparentemente, essa estrutura considera apenas os dados de um período como entrada e outro como saída. No entanto, ela é construída de forma que as informações de dois ou mais períodos sejam também consideradas neste modelo. Ou seja, os dados de diferentes períodos são armazenados como se eles estivessem na mesma coluna de uma planilha, formando blocos. Por exemplo, considerando-se inicialmente os dados de 1960 e 1970, o primeiro bloco é organizado de tal forma que os dados de 1960 sejam utilizados como entrada e 1970 como dado de saída. No segundo bloco, toma-se os dados de 1970 como entrada e 1980 como saída, organizando-os abaixo dos dados de 1960 e 1970, respectivamente. Esse procedimento é realizado conforme a disponibilidade de dados. No estudo de Manzato e Rodrigues da Silva (2010), três blocos foram considerados para a fase de estimativa de parâmetros, sendo o último bloco composto pelos anos de 1980 como entrada e 1990 como saída. O ano de 2000 foi preservado para a fase de validação. Sendo assim, se a primeira estrutura tem “ $w$ ” observações (ou municípios, neste caso), a segunda teve “ $3w$ ” observações. Com a inclusão dos dados de 2010, foi possível aumentar esse número para “ $4w$ ” observações, considerando que a fase de estimativa de parâmetros pôde contar com um período adicional. Ou seja, o quarto e último bloco foi composto pelos anos de 1990 como entrada e 2000 como saída. Isso proporcionou mais observações disponíveis à fase de estimativa de parâmetros. Dessa forma, essa segunda estrutura passou a ser chamada de “T, T + 10 com 4 períodos”. A Figura 1 ilustra o esquema das estruturas dos modelos espaciais já adaptados a partir de Manzato e Rodrigues da Silva (2010).



**Figura 1.** Esquema das estruturas dos modelos espaciais, adaptados de Manzato e Rodrigues da Silva (2010)

Considerando ambas as estruturas de modelos e as variáveis definidas anteriormente, Manzato e Rodrigues da Silva (2010) e Ajauskas *et al.* (2012) organizaram uma série de seis modelos, divididos em três grupos, conforme ilustrado na Tabela 1. Cabe ressaltar que a referida tabela já incorpora a nova nomenclatura para a estrutura dos modelos desenvolvidos no presente estudo.

No primeiro grupo, os modelos 1 e 2 levam em conta apenas as variáveis relacionadas à densidade populacional. Ou seja, embora a ideia inicial fosse uma composição de indicadores populacionais e de oferta de infraestrutura rodoviária, esse grupo de modelos proporciona uma análise da definição de RUHs, por meio da metodologia proposta, desconsiderando-se a influência da infraestrutura de transportes. Isso serve para avaliar o impacto da incorporação desta infraestrutura nos modelos espaciais

conforme previsto no segundo grupo de modelos. Dessa forma, os modelos 3 e 4 são compostos por variáveis relacionadas à densidade populacional e ao índice de oferta rodoviária. Convém ressaltar que o referido índice é incorporado aos modelos por meio do cálculo de sua variação observada entre dois períodos em análise. Por exemplo, no modelo de estrutura “4 IN 1 OUT”, a variação desse índice entre os períodos “ $t - 10$ ” e “ $t - 40$ ” foi utilizada. Essa variação foi determinada pela relação “ $t - 10$ ” / “ $t - 40$ ”. Para o modelo de estrutura “T, T+10 com 4 períodos”, essa variação se referiu a dois anos consecutivos entre os respectivos pares de períodos analisados. Em ambos os casos, tal variação foi ainda normalizada entre 0,1 e 0,9 para evitar valores nulos ou extremamente altos, o que poderia produzir inconsistências durante a modelagem. Por fim, no terceiro grupo de modelos (5 e 6), elaborados particularmente por Ajauskas *et al.* (2012), buscou-se uma simplificação dos modelos anteriores por meio da utilização de apenas dois atributos de entrada e um de saída. Como dados de entrada foram utilizados o quadrante do gráfico de Moran obtido para densidade populacional e o quadrante do gráfico de Moran obtido para o índice de oferta rodoviária (respectivamente,  $QPD_x$  e  $QIC_x$ ). Como dado de saída, os modelos fornecem o valor do quadrante do gráfico de Moran referente à densidade populacional ( $QPD_x$ ).

**Tabela 1:** Resumo dos modelos espaciais em função de sua estrutura e variáveis consideradas

Modelo	Grupo	Estrutura	Variáveis de Entrada	Variável de Saída
1	1	4 IN 1 OUT	$PDX, QPD_x, PDj, n\_HH_x^{PDX}, n\_LL_x^{PDX}, n\_LH_x^{PDX}$ e $n\_HL_x^{PDX}$	$PDX$
2	1	T, T+10 com 4 períodos	$PDX, QPD_x, PDj, n\_HH_x^{PDX}, n\_LL_x^{PDX}, n\_LH_x^{PDX}$ e $n\_HL_x^{PDX}$	$PDX$
3	2	4 IN 1 OUT	$PDX, QPD_x, PDj, n\_HH_x^{PDX}, n\_LL_x^{PDX}, n\_LH_x^{PDX}, n\_HL_x^{PDX}$ e $\Delta IC_x$	$PDX$
4	2	T, T+10 com 4 períodos	$PDX, QPD_x, PDj, n\_HH_x^{PDX}, n\_LL_x^{PDX}, n\_LH_x^{PDX}, n\_HL_x^{PDX}$ e $\Delta IC_x$	$PDX$
5	3	4 IN 1 OUT	$QPD_x, QIC_x$	$QPD_x$
6	3	T, T+10 com 4 períodos	$QPD_x, QIC_x$	$QPD_x$

Por fim, cabe ressaltar que os modelos espaciais foram aqui desenvolvidos utilizando-se uma estrutura de redes neurais artificiais (RNA) com o auxílio do *software* EasyNN-Plus. Para tanto, na fase de estimativa de parâmetros, as observações foram subdivididas aleatoriamente em um conjunto de treinamento e um de validação interna (requerida pela modelagem com RNA). Esses conjuntos corresponderam, respectivamente, a 70% e 30% das observações. Na fase de validação dos parâmetros, 100% das observações foram utilizadas. Além disso, o *software* de sistema de informação geográfica (SIG) Maptitude foi utilizado para a manipulação dos dados e geração de mapas temáticos e o *software* GeoDa serviu para a aplicação da técnica ESDA.

### 3. RESULTADOS

Na Tabela 2 estão apresentados os resultados dos desempenhos dos modelos espaciais desenvolvidos. Tais desempenhos foram obtidos em função da porcentagem de acertos de cada modelo, calculando-se a relação entre o dado estimado e o real. Esses desempenhos são ainda fornecidos tanto em termos globais de cada modelo, como para os parciais em cada quadrante. Cabe ressaltar que, para fins de padronização desta análise, os resultados dos modelos 1 a 4, cuja variável de saída é a densidade populacional, foram submetidos à técnica ESDA para a obtenção do respectivo quadrante. Isto é, a porcentagem de acertos é dada pela relação entre o quadrante estimado e o quadrante real. Essa análise é efetuada com os dados da fase de validação dos parâmetros, utilizando-se o ano de 2010.

Conforme os resultados globais mostrados na Tabela 2, percebe-se que os modelos de estrutura “T, T + 10 com 4 períodos” (2,4 e 6) apresentaram em geral um desempenho ligeiramente superior àqueles

de estrutura "4 IN 1 OUT" (1, 3 e 5), considerando-se os pares de modelos dentro de cada grupo. Mais especificamente, é possível observar que o modelo 2 apresentou o melhor desempenho global, seguido do modelo 6, e por fim o modelo 4 com o menor desempenho.

**Tabela 2:** Porcentagem de acertos parciais e globais dos modelos espaciais

Modelo	Grupo	HH	LL	LH	HL	Global
1	1	91,4	98,7	100	87,5	97,4
2	1	91,4	98,9	100	93,8	98,3
3	2	86,2	98,2	96,0	81,3	96,9
4	2	93,1	98,7	100	81,3	96,9
5	3	94,8	99,5	92,0	93,8	97,8
6	3	94,8	99,8	84,0	93,8	98,1

No entanto, analisando-se os modelos de estrutura "T, T+10 com 4 períodos" que estimam a densidade populacional (2 e 4), a porcentagem de acertos no quadrante HH foi maior para o modelo 4, o que proporciona um destaque para este modelo em relação ao modelo 2. Esse argumento se baseia na ideia de que RUHs são definidas sobretudo por meio do quadrante HH segundo a metodologia aqui aplicada. Cabe, portanto, ressaltar a importância da inclusão dos dados de oferta rodoviária em proporcionar uma melhoria no desempenho dos modelos. Embora essa melhoria pareça pequena, o fato é que a combinação de indicadores populacionais e de oferta de infraestrutura de transportes se confirma como uma boa estratégia para a definição de RUHs, conforme defendido por Manzato e Rodrigues da Silva (2010).

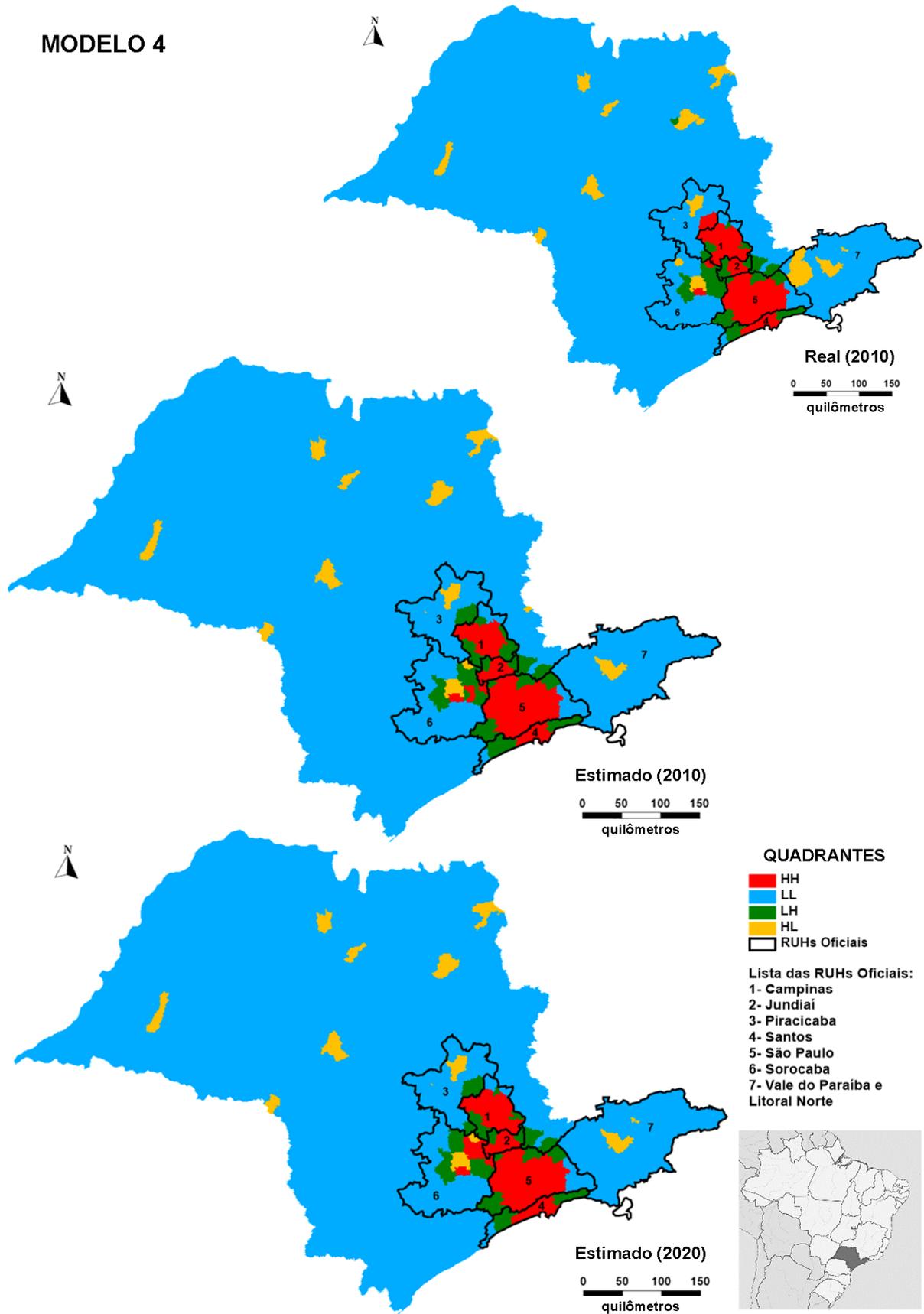
Tal fato também pode ser observado nos resultados do modelo 6, destacando-se que a porcentagem de acertos no quadrante HH é a maior dentre todos os modelos. A única desvantagem deste modelo é a questão de não ser possível estimar a densidade populacional, mas apenas os seus quadrantes diretamente. No entanto, trata-se de um modelo que também considera a relação entre oferta de transportes e distribuição populacional e o seu bom desempenho apresentado neste estudo está de acordo com o que Ajauskas *et al.* (2012) obtiveram.

Outra questão que merece destaque retoma os dois tipos de estruturas concebidas para os modelos e sua relação com a definição de RUHs. Conforme argumentado, a estrutura "T, T + 10 com 4 períodos" proporcionou modelos com desempenhos ligeiramente superiores. Em princípio, esses modelos aparentemente estariam baseados em dados de apenas uma década. No entanto, conforme descrito na metodologia, essa estrutura incorpora dados de um período mais longo do que dez anos, organizando pares de décadas em blocos verticais durante a fase de estimativa de parâmetros. Dessa maneira, pode-se dizer que essa estrutura também relaciona o comportamento histórico dos dados.

Além disso, o fato do processo de modelagem com RNAs ser baseado em uma técnica de mineração de dados, a estrutura que melhor acomodou os dados para o caso aqui estudado foi a "T, T + 10 com 4 períodos". No entanto, a estrutura 4 IN 1 OUT tem sua relevância, considerando que o desempenho dos modelos nela baseados também apresentaram bons resultados. Recomenda-se, portanto, continuar testando o seu desempenho em aplicações futuras que estejam baseadas na metodologia aqui apresentada. Dessa maneira, pode-se avaliar a significância dos resultados dos vários tipos de modelos e suas respectivas estruturas de forma comparativa.

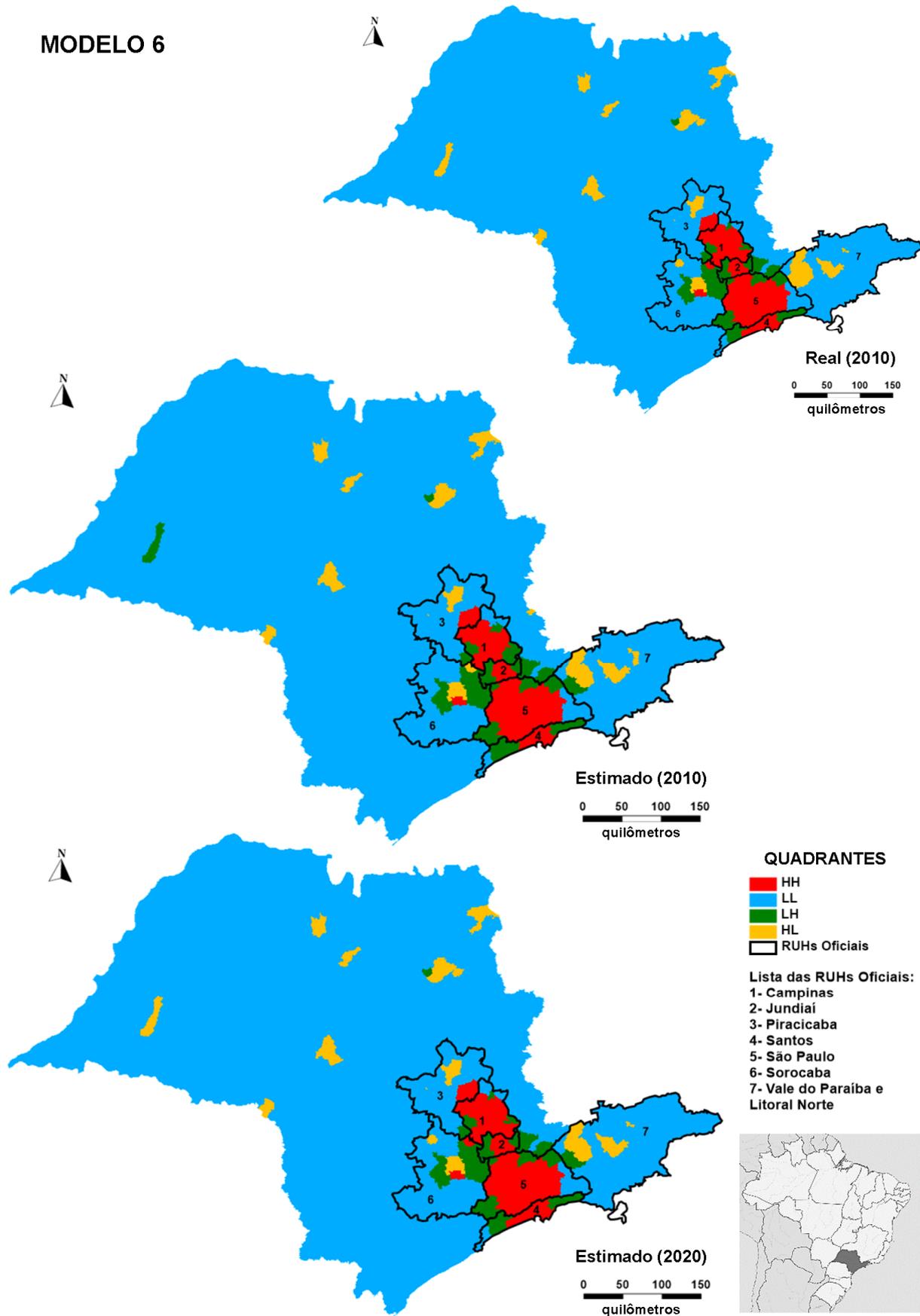
Em conjunto com os desempenhos dos modelos, é importante também analisar a distribuição espacial dos resultados, os quais estão ilustrados nas Figuras 2 e 3. Respectivamente, essas figuras se referem aos modelos 4 e 6, por eles terem proporcionado os melhores desempenhos conforme discutido anteriormente. Em cada figura estão apresentados os dados de densidade populacional real e a estimada pelos modelos para o ano de 2010 (fase de validação) por meio de *Box Maps*. Além disso, em cada *Box Map* foi adicionado o contorno das RUHs oficiais do estado de São Paulo para sua comparação com os resultados da metodologia aqui empregada. A análise visual das Figuras 2 e 3 permite observar que os modelos 4 e 6 de fato apresentam resultados consistentes, sobretudo considerando a distribuição dos municípios no quadrante HH.

**MODELO 4**



**Figura 2.** Box maps representando a densidade populacional real e a estimada para os dados de 2010, além da previsão da densidade populacional para 2020, fornecidas pelo modelo 4

**MODELO 6**



**Figura 3.** Box maps representando a densidade populacional real e a estimada para os dados de 2010, além da previsão da densidade populacional para 2020, fornecidas pelo modelo 6

Uma análise adicional ainda é apresentada nas Figuras 2 e 3, em que uma estimativa da densidade populacional para 2020 foi obtida segundo os modelos espaciais destacados. Essa análise tem um caráter prospectivo, verificando-se o potencial de previsão dos modelos. Nesse caso, por meio de uma comparação visual, visualiza-se variações pontuais de alguns municípios que “evoluiriam” do quadrante LL para o quadrante LH e do quadrante LH para o quadrante HH. Essas variações podem ser observadas sobretudo nas RUHs oficiais criadas recentemente, como é o caso das RUHs de Sorocaba e do Vale do Paraíba e Litoral Norte. No entanto, estudos mais aprofundados sobre esses processos serão apresentados oportunamente em trabalhos futuros.

#### 4. CONCLUSÕES

Em função dos resultados apresentados neste trabalho, cujo objetivo foi incorporar os dados demográficos e de oferta de infraestrutura rodoviária do ano de 2010 promovendo uma atualização dos modelos espaciais propostos por Manzato e Rodrigues da Silva (2010) e Ajauskas *et al.* (2012), algumas conclusões relevantes foram obtidas. A nova estimativa dos parâmetros com os dados de 2010 resultou em modelos espaciais cujos desempenhos foram bastante semelhantes aos apresentados pelos referidos autores. Em particular, os modelos que incorporam dados de oferta rodoviária apresentaram resultados superiores, sobretudo no quadrante HH, ao qual estão associadas as RUHs conforme a metodologia aqui utilizada. Dessa maneira, tais resultados confirmam a proposta metodológica de se empregar uma composição de indicadores de oferta de infraestrutura de transportes e populacionais baseada em recursos de análise espacial para a definição de RUHs.

Dentre os modelos desenvolvidos, dois se destacaram: o modelo 4 e o modelo 6. Este último tem a vantagem de ser elaborado com um reduzido número de variáveis. No entanto, o referido modelo não estima o valor da densidade populacional, mas o quadrante do gráfico de Moran referente à densidade populacional. Isso pode ser uma desvantagem em alguns casos quando o valor da densidade populacional for necessário. Nesse sentido, o modelo 4 por sua vez, tem a capacidade de contornar essa desvantagem, pois a variável estimada neste modelo refere-se à densidade populacional. Este modelo apresentou um desempenho inferior ao modelo 6 considerando-se a análise da porcentagem de acertos no quadrante HH e o seu desempenho global foi o menor de todos os modelos desenvolvidos. No entanto, considerando-se que a incorporação da infraestrutura rodoviária na modelagem ainda é relevante para a análise da definição de RUHs, assume-se que o valor global de 96,9% de acertos seja bastante razoável.

Em resumo, as análises desenvolvidas neste estudo demonstram novamente o potencial da aplicação das técnicas de estatística espacial e modelagem espacial como uma abordagem promissora para a definição e monitoramento de RUHs. Em particular, destaca-se nessa abordagem a utilização de dados relativamente fáceis de serem obtidos e manipulados para tal finalidade. Embora o estudo de caso aqui desenvolvido contemple apenas o estado de São Paulo, análises semelhantes podem ser aplicadas em outras regiões do país, como mostrado por Rodrigues da Silva *et al.* (2014). É evidente que na maioria dos casos o emprego de dados de oferta rodoviária contemple o principal (senão o único) modo de transporte. Mas a metodologia permite incorporar outros modos e adaptá-la à realidade da região que se deseja analisar. Um exemplo interessante seria o caso da região Norte do Brasil, onde o transporte fluvial é bastante importante.

#### AGRADECIMENTOS

Os autores agradecem às agências FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo – Processos 2014/23967-7 e 2014/25424-0), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico – processos 471397/2013-0 e 302177/2015-9) e CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo apoio concedido em diferentes fases da pesquisa que deu origem a este trabalho. Os autores agradecem também à *Caliper Corporation* pela doação da licença educacional do *software* Maptitude e ao DER-SP pelos dados utilizados neste estudo.

#### REFERÊNCIAS

Aguiar, L. L. e G. G. Manzato (2017) Análise exploratória de dados de viagens pendulares visando a definição de Regiões Urbanas Homogêneas. *Transportes*, no prelo.

- 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 15<sup>th</sup> International Conference on Computers in Urban Planning and Urban Management (CUPUM)*, Adelaide, Austrália.
- 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.
- 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
- Batty, M. e Y. Xie (1994) From cells to cities. *Environment and Planning B: Planning and Design*, v. 21, p. s31-s48. DOI:10.1068/b21S031
- Batty, M.; H. Couclelis e M. Eichen (1997) Urban systems as cellular automata. *Environment and Planning B: Planning and Design*, v. 24, p. 159-164. DOI:10.1068/b240159
- Baum-Snow, N. (2010) Changes in Transportation Infrastructure and Commuting Patterns in US Metropolitan Areas, 1960-2000. *The American Economic Review: Papers & Proceedings*, v. 100, n. 2, p. 378-382. DOI:10.1257/aer.100.2.378
- 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
- Breitung, W. (2011) Borders and the City: Intra-Urban Boundaries in Guangzhou (China). *Quaestiones Geographicae*, v. 30, n. 4, p. 55-61. DOI:10.2478/v10117-011-0038-5
- Cecchini, A. (1996) Urban modelling by means of cellular automata: generalised urban automata with the help on-line (AUGH) model. *Environment and Planning B: Planning and Design*, v. 23, p. 721-732. DOI:10.1068/b230721
- Cheshire, P. C. e D. G. Hay (1989) *Urban problems in Western Europe: an economic analysis*. Unwin Hyman, London.
- Clarke, K. C.; S. Hoppen e L. Gaydos (1997) A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B: Planning and Design*, v. 24, p. 247-261. DOI:10.1068/b240247
- Coomes, 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.
- Debbage, N.; B. Bereitschaft e J. M. Shepherd (2017) Quantifying the Spatiotemporal Trends of Urban Sprawl Among Large U.S. Metropolitan Areas via Spatial Metrics. *Applied Spatial Analysis and Policy*, v. 10, p. 317-345. DOI:10.1007/s12061-016-9190-6
- Dias, R. S.; G. G. Manzato e A. N. Rodrigues da Silva (2014) A capacidade de infraestrutura rodoviária e sua relação com o processo de metropolização. *Anais do XXVIII Congresso de Pesquisa e Ensino em Transportes*, Curitiba, Brasil.
- Gouvêa, R. G. (2005) *A questão metropolitana no Brasil*. FGV, Rio de Janeiro.
- 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
- IBGE (2010) Instituto Brasileiro de Geografia e Estatística. Censo 2010. Disponível em: <http://www.censo2010.ibge.gov.br/>
- Kauffmann, A. (2012) *Delineation of city regions based on commuting interrelations: The example of large cities in Germany*. IWH-Diskussionspapiere, n. 2012,4. Disponível em: <http://nbn-resolving.de/urn:nbn:de:101:1-201212176768>.
- Kourtit, K. e P. Nijkamp (2013). In praise of megacities in a global world. *Regional Science Policy and Practice*, v. 5, p. 167-182. DOI:10.1111/rsp3.12002
- Kourtit, 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
- Lin, G. C. S. (1999) Transportation and Metropolitan Development in China's Pearl River Delta: The Experience of Panyu. *Habitat International*, v. 23, n. 2, p. 249-270. DOI:10.1016/S0197-3975(98)00049-6
- Magalhães, M. T. Q.; G. L. Teixeira e Y. Yamashita (2004) Indicadores de Cobertura Espacial para Diagnóstico da Dotação de Estrutura de Circulação Rodoviária Brasileira. *Anais do XVIII Congresso de Pesquisa e Ensino em Transportes*, Florianópolis, Brasil, v. 2, p. 880-891.
- Manzato, G. G. e A. N. Rodrigues da Silva (2006) Incorporando um indicador de oferta de infraestrutura de transportes na definição de regiões metropolitanas. *Anais do XX Congresso de Pesquisa e Ensino em Transportes*, Brasília, Brasil, v. 1, p. 341-352.
- Manzato, G. G. e A. N. Rodrigues da Silva (2007) Uma estrutura conceitual para a definição de regiões urbanas homogêneas. *Anais do XXI Congresso de Pesquisa e Ensino em Transportes*, Rio de Janeiro, Brasil.

- 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.; A. J. Soares e A. N. Rodrigues da Silva (2006) Aglomerações urbanas e oferta de transportes no estado de São Paulo, Brasil. *Anais do 2º Congresso Luso-Brasileiro para o Planejamento Urbano Regional Integrado Sustentável*, Braga, Portugal.
- 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*, Foz do Iguaçu, Brasil.
- 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
- 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 (2000) Standards for defining metropolitan and micropolitan statistical areas. *Federal Register*, v. 65, n. 249, December 27, 2000.
- 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
- Rodrigues da Silva, A. N.; M. S. Costa e M. H. Macedo (2008) Multiple Views of Sustainable Urban Mobility: The Case of Brazil. *Transport Policy*, v. 15, n. 6, p. 350-360. DOI: 10.1016/j.tranpol.2008.12.003
- Rodríguez-Pose, A. e T. D. Ketterer (2012) Do local amenities affect the appeal of regions in Europe for migrants? *Journal of Regional Science*, v. 52, p. 535-561. DOI:10.1111/j.1467-9787.2012.00779.x
- 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](http://www.legislacao.sp.gov.br/legislacao/index.htm).
- UN (2007) United Nations, Population fund, state of world population 2007, New York: UN.
- UN (2015) United Nations, Department of Economic and Social Affairs, Population Division. World Urbanization Prospects: The 2014 Revision, (ST/ESA/SER.A/366).
- 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.
- White, R. e G. Engelen (1993a) Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns. *Environment and Planning A*, v. 25, p. 1175-1199. DOI:10.1068/a251175
- White, R. e G. Engelen (1993b) Cellular dynamics and GIS: modelling spatial complexity. *Geographical Systems*, v. 1, p. 237-253. DOI:10.2307/1587033
- 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