

ESCOLHA DA ESCALA PARA MODELAGEM GEOESTATÍSTICA DE VARIÁVEIS DE DEMANDA POR TRANSPORTES

Samille Santos Rocha

Cira Souza Pitombo

Universidade de São Paulo

Escola de Engenharia de São Carlos

Luís Henrique Magalhães Costa

Universidade Estadual Vale do Acaraú - UVA

RESUMO

Dados de demanda por transportes não são independentes no espaço. Assim, a utilização de técnicas que consideram a autocorrelação espacial, como a geoestatística, pode aprimorar as estimativas. Em demanda por transportes, a escolha da escala de representação de uma variável é essencial para a modelagem geoestatística, pois é necessário considerar dois fatores: 1) a representação do comportamento individual (domicílios), através da sua distribuição espacial e, 2) a qualidade dos modelos espaciais, comprometida na abordagem desagregada, devido à alta aleatoriedade espacial. Desta forma, este artigo preenche uma lacuna de pesquisa: a busca pela escala apropriada para representação de dados de demanda por transportes para aplicação de modelagem geoestatística. Assim, este artigo apresenta a implementação de um método, automatizado, para a escolha da dimensão de uma grade regular para agregar dados, por comparação de distribuições populacionais através de testes não paramétricos. Finalmente, foi possível determinar o limite de dimensão de uma grade regular que considerasse os fatores mencionados anteriormente.

ABSTRACT

Travel demand data is not independent in space. Thus, using techniques that consider spatial autocorrelation, such as geostatistics, can improve estimates. In travel demand, the choice of the representation scale of a variable is essential for geostatistical modeling as two factors need to be considered: 1) the representation of individual behavior (households), through its spatial distribution and 2) the quality of spatial models, compromised in the disaggregated approach, due to high spatial randomness. Thus, this paper fills a research gap: searching for the appropriate scale to represent travel demand data for geostatistical modeling applications. This article presents an automated method implementation to choose the size of a regular grid for data aggregation by comparing population distributions using non-parametric tests. Finally, the size limit of a regular grid was determined considering the factors mentioned previously.

1. INTRODUÇÃO

A análise de demanda por transportes relaciona-se, também, com a localização espacial dos indivíduos e/ou domicílios e dos destinos das viagens. Modelos que utilizam a localização geográfica das variáveis são ferramentas importantes para o aprimoramento das estimativas de dados relacionados à demanda por transportes (Eom *et al.*, 2006; Pitombo *et al.*, 2015; Pitombo *et al.*, 2010; Klatko *et al.*, 2017).

Dentre as técnicas que consideram a dependência espacial de variáveis, destaca-se a geoestatística. A geoestatística é um método de análise que possibilita o desenvolvimento de estudos onde os valores das variáveis não são independentes no espaço. A aplicação da geoestatística permite, principalmente, prever valores de uma variável em pontos não amostrados, além de ser possível avaliar a acurácia das estimativas. A capacidade de prever valores de uma variável em locais sem informações, através de pontos conhecidos, é uma vantagem da geoestatística em relação a outros métodos de análise espacial (Matheron, 1970; Isaaks e Srivastava, 1989). No entanto, é fundamental considerar que dados relacionados à demanda por transportes possuem algumas particularidades quando analisados através da geoestatística. As unidades de área que, geralmente, representam dados relacionados à demanda por transportes, como zonas de tráfego e setores censitários, por exemplo, possuem formas e dimensões irregulares (Goovaerts, 2008). Este é mais um obstáculo da aplicação da geoestatística em tais estudos, visto que esta técnica foi desenvolvida para ser aplicada a

unidades de áreas regulares (com mesma forma e tamanho). Por isso, a escala de amostragem é um dos fatores que influenciam na qualidade dos modelos.

Dados desagregados, apesar de ser a maneira mais indicada para análises tradicionais de demanda por transportes, geralmente, apresentam alta variabilidade a curtas distâncias. Essa característica dificulta a modelagem do comportamento espacial referente a indivíduos ou domicílios (Lindner, 2015; Marques e Pitombo, 2018; Marques, 2019; Rocha, 2019).

A escolha da escala de representação espacial dos dados de uma variável é imprescindível para a modelagem geoestatística de variáveis relacionadas à demanda por transportes, pois há duas importantes questões a serem consideradas: 1) representar adequadamente o comportamento individual, através da sua distribuição espacial que associa os pontos à localização domiciliar e, 2) prezar pela qualidade dos modelos espaciais, comprometida na abordagem desagregada, devido à alta aleatoriedade espacial.

Este artigo procura discutir o desafio de representar adequadamente dados referentes a viagens domiciliares, apresentando-o como um problema metodológico inerente à modelagem geoestatística. Para isso, será proposto um método, automatizado, para a definição da escala de análise a fim de realizar a modelagem de dados relativos a viagens. Portanto, este trabalho visa à busca de suporte homogêneo, para representar dados desagregados, a fim de diminuir a aleatoriedade espacial, associada aos domicílios, e solucionar o problema de irregularidade de suporte, associado às zonas de tráfego.

2. O PROBLEMA DA ESCALA EM ANÁLISE ESPACIAL

O conceito de escala não é apenas utilizado na Geografia, mas em vários campos de estudo, devido à sua influência no processo de experimentação, sendo, portanto, uma medida que dá visibilidade ao fenômeno (Castro *et al.*, 1995). Desta forma, a escala define o nível de análise, ou recorte do objeto de estudo.

Geralmente, dados espaciais são obtidos pontualmente ou representados através de uma unidade de área (Gelfand *et al.*, 2001). Em estudo de dados agregados por área, o principal problema é a definição de critérios para definição dos polígonos, uma vez que diferentes agrupamentos podem gerar análises com diferentes resultados.

Por motivos como confidencialidade, por exemplo, na maioria das pesquisas sobre informações socioeconômicas da população, os dados coletados são vinculados a áreas. No entanto, o que se deseja conhecer, na maioria das pesquisas, são características relacionadas a indivíduos (dados desagregados). Logo, os possíveis efeitos da escala não podem ser ignorados, pois podem ocorrer possíveis erros na interpretação dos resultados. Por isso, em modelagem de dados espaciais, problemas relacionados à escala teriam como solução mais adequada a manutenção da mesma escala e características dos agrupamentos para a obtenção de dados amostrais (Openshaw, 1976). Em análise espacial, fenômenos relacionados à escala vêm sendo bastante discutidos. Um deles é a existência do Problema da Unidade de Área Modificável (*Modifiable Areal Unit Problem – MAUP*) que pode ocorrer devido ao nível de agregação dos dados, através da modificação da escala ou do nível de agregação de polígonos.

2.1. O Problema da Unidade de Área Modificável – MAUP

Análises de mudança de escala, com base na mudança de tamanhos das unidades de área ou polígonos, já foram bastante discutidas em análise espacial através do conceito do Problema da Unidade de Área Modificável. O efeito de escala está relacionado à influência do nível de agregação de variáveis espaciais, onde os resultados tendem a variar de acordo o nível de

resolução espacial (Páez e Scott, 2005). O *MAUP* envolve duas diferentes questões: a) a relação com a localização dos dados e, b) sua adequada representação, que depende do sistema de zoneamento adotado (Fotheringham, 1989).

Em estudos urbanos, geralmente, os recortes espaciais não levam em conta fatores socioeconômicos da população. Desta forma, a mudança da escala de análise pode interferir nas características estatísticas de uma amostra e na variabilidade dos resultados. Em áreas com desigualdades socioeconômicas, como no caso de muitos municípios brasileiros, dados agregados em uma mesma área ou polígono podem estar representando diferentes grupos sociais. Estatisticamente, valores médios por áreas podem divergir essencialmente de dados desagregados ou individuais (Holt *et al.*, 1996). Como a análise por unidade de áreas, muitas vezes, pretende inferir dados pontuais com base em médias ou taxas de uma variável nos blocos ou zonas, o comportamento individual, em certas áreas, pode ser bem diferente da média encontrada em análises agregadas, podendo ocorrer erros grosseiros de interpretação.

Desta maneira, diferentes níveis de agregação podem levar a incertezas na interpretação dos resultados, levando ao problema da falácia ecológica, conceito associado ao *MAUP*. A falácia ecológica ocorre quando as análises baseadas em dados agregados levam a conclusões diferentes das análises baseadas em dados individuais (Gotway e Young, 2002). No entanto, é importante considerar que sua existência depende da natureza da agregação utilizada e do grau de homogeneidade dos agrupamentos (Openshaw, 1984).

Em geoestatística, o *MAUP* já vem sendo discutido em alguns trabalhos. Em mineração, por exemplo, amostras com apenas alguns metros cúbicos de volume são utilizadas para estimar o grau médio de minério em unidades de volume bem maiores. Cressie (1996) identifica esse problema na mineração como um caso especial de problema de mudança de suporte – *Change of Support Problem (COSP)*. Dado que a geoestatística busca realizar a inferência de informações em um domínio espacial contínuo, não considerar a mudança de suporte espacial pode resultar em previsões altamente tendenciosas.

A partir da relevância da mudança de escala para a geoestatística, este estudo propõe a busca da escala apropriada para variáveis de demanda por transportes, considerando a distribuição populacional dos dados mais desagregados, além de homogeneidade de suporte. Sabe-se que resultados totalmente confiáveis, onde as soluções das análises não dependam do sistema de zoneamento utilizado, são assegurados a partir de dados individuais ou domiciliares. No entanto, estes tipos de dados, relacionados à demanda por transportes, quando disponíveis, apresentam elevado grau de aleatoriedade espacial, resultando em modelos geoestatísticos com pouca estrutura espacial.

3. GEOESTATÍSTICA: CONCEITOS BÁSICOS

A geoestatística pode ser definida como a aplicação a Teoria das Variáveis Regionalizadas (VR) para a compreensão da distribuição espacial de dados (Journel, 1986). A teoria geoestatística associa a aleatoriedade do fenômeno usando um modelo estocástico, onde uma VR é considerada como sendo uma entre muitas realizações possíveis de uma função aleatória (Matheron, 1970; Chilès e Delfiner, 2012). Para ser considerada uma VR, um conjunto amostral, necessariamente, apresenta uma forte dependência em relação à sua localização espacial (Matheron, 1970).

Em geoestatística, a variabilidade entre dois pontos, separados por uma distância h , pode ser expressa pela função do semivariograma (Equação 1), representada por $\gamma(x, h)$.

$$y(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \quad (1)$$

Sendo,

$\gamma(h)$ a semivariância para a distância h , que separa as amostras; $Z(x_i)$ o valor de uma VR localizada em i ; $Z(x_i + h)$ o valor de uma variável regionalizada localizada em $i + h$; $N(h)$ o número de pares de pontos encontrados através de h .

A variabilidade do fenômeno é então representada graficamente pelo semivariograma experimental. Em seguida, a modelagem da variabilidade dos dados é feita através de uma curva de ajuste (semivariograma teórico). A Figura 1 apresenta os parâmetros do semivariograma teórico, que serão utilizados na etapa posterior relativa à estimativa, conhecida como krigagem.

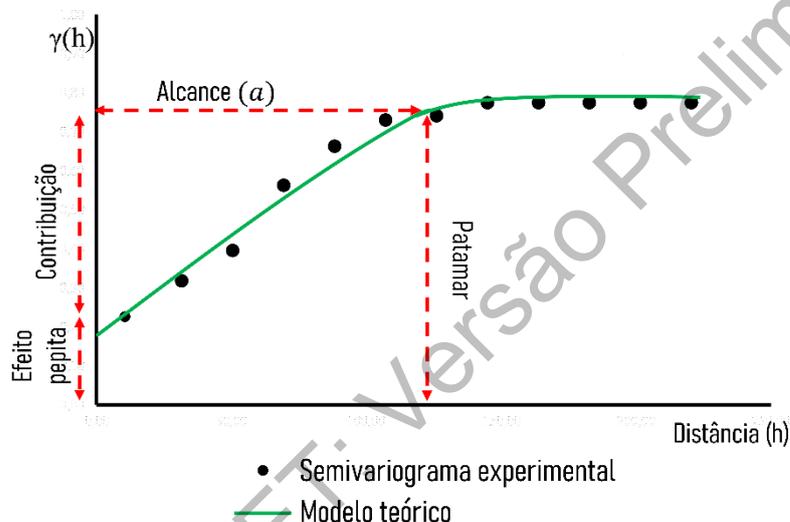


Figura 1: Estrutura do semivariograma teórico. Fonte: Adaptado de Druck *et al.* (2004).

O semivariograma expressa variabilidade em função da distância e direção, além do comportamento da variabilidade a pequenas distâncias próximo à origem (Yamamoto, 2001). A seguir são definidos os parâmetros apresentados na Figura 1.

- **Efeito Pepita (C_0):** uma aparente descontinuidade junto à origem, onde a curva do modelo toca o eixo γ . Um valor alto deste parâmetro, também pode indicar a presença de grande variabilidade dos dados a curtas distâncias, o que significa dizer que os dados possuem alta aleatoriedade espacial; **Alcance (α):** corresponde à distância máxima em que as amostras apresentam correlação espacial. A partir de um alcance α , o patamar é atingido (Chilès e Delfiner, 2012; Druck *et al.*, 2004; Journel e Huijbregts, 1978); **Patamar ($C = C_0 + C_1$):** valor de γ máximo da curva correspondente à amplitude. C_1 é a contribuição ou variância espacial do modelo. A partir do patamar a variância estabiliza-se a uma distância α e considera-se que não há mais dependência espacial, porque a variância entre as amostras torna-se, aproximadamente, constante (Chilès e Delfiner, 2012; Druck *et al.*, 2004).

Os parâmetros, apresentados pela Figura 1, são utilizados na etapa de estimativa geoestatística, a partir da krigagem. A krigagem é um algoritmo que realiza a interpolação espacial de dados através dos pesos associados a cada amostra, atribuídos a partir dos parâmetros do semivariograma teórico (Chilès e Delfiner, 2012; Wackernagel, 2003). Quanto mais próximo uma amostra estiver de um ponto a ser estimado, maior será sua influência no valor deste ponto. Por outro lado, quanto mais agrupados estiverem os dados, menor será o peso individual de

cada amostra no sistema de estimação de um ponto (Soares, 2000).

Na krigagem, a interpolação é realizada estabelecendo uma grade regularmente espaçada, onde cada nó da grade corresponde a um ponto x_0 a ser estimado. Como resultado da interpolação, tem-se uma superfície contínua de valores estimados de $Z(x)$. Além disso, como método estatístico, a krigagem fornece medidas de incerteza como, por exemplo, o erro e a variância da estimação (Wackernagel, 2003). Na Krigagem Ordinária, estimador utilizado neste trabalho, o valor desconhecido $Z^*(x_0)$ (Equação 2) pode ser estimado a partir de uma combinação linear dos pesos λ_i com n valores vizinhos $Z(x_i)$ do ponto a ser estimado (Journel, 1989).

$$Z_{x_0}^* = \sum_{i=1}^n \lambda_i Z(x_i) \quad (2)$$

Os pesos λ_i são calculados de modo a garantir que o estimador seja não-enviesado e a variância de estimação seja mínima. Por isso, o estimador KO é bastante utilizado. Os pesos ótimos e variância mínima são obtidos pelos parâmetros de Lagrange (Journel e Huijbregts, 1978). A esperança de um erro mínimo (Equação 3), é dada pela condição de não-enviesamento (Equação 4):

$$E[Z_{x_0} - Z_{x_0}^*] = 0 \quad (3)$$

$$\sum_{i=1}^n \lambda_i = 1 \quad (4)$$

Atendendo a esta condição, a variância das estimativas (Equação 6) será minimizada. Tal restrição é obtida construindo-se um sistema de n equações com n incógnitas, então passa-se a $n + 1$ equações, introduzindo uma incógnita, também desconhecida, para satisfazer a condição de variância mínima estabelecida pela Krigagem Ordinária. Esta incógnita é o multiplicador de Lagrange (μ) (Landim, 2006). Logo, a partir do sistema de equações da krigagem ($n + 1$), são calculados os ponderadores λ_n e a variância da estimação:

$$\begin{cases} \sum_{i=1}^n \lambda_i \gamma(x_i - x_j) + \mu = \gamma(x_0 - x_j), i = 1, \dots, n \\ \sum_{j=1}^n \lambda_j = 1 \end{cases} \quad (5)$$

$$\sigma_{KO}^2(x_0) = \sum_{i=1}^n \lambda_i \gamma(x_0 - x_i) + \mu \quad (6)$$

O sistema de equações da Krigagem Ordinária pode ser representado de forma matricial,

$$[K] \cdot [\lambda] = [M] \quad (7)$$

onde,

K é a matriz da variância entre as amostras; λ_n são os ponderadores da krigagem e M é matriz da variância entre as amostras e o ponto a ser previsto. Para obter os valores dos ponderadores $[\lambda]$, são obtidos os produtos entre a matriz inversa de K e o segundo membro da matriz (Journel e Huijbregts, 1978).

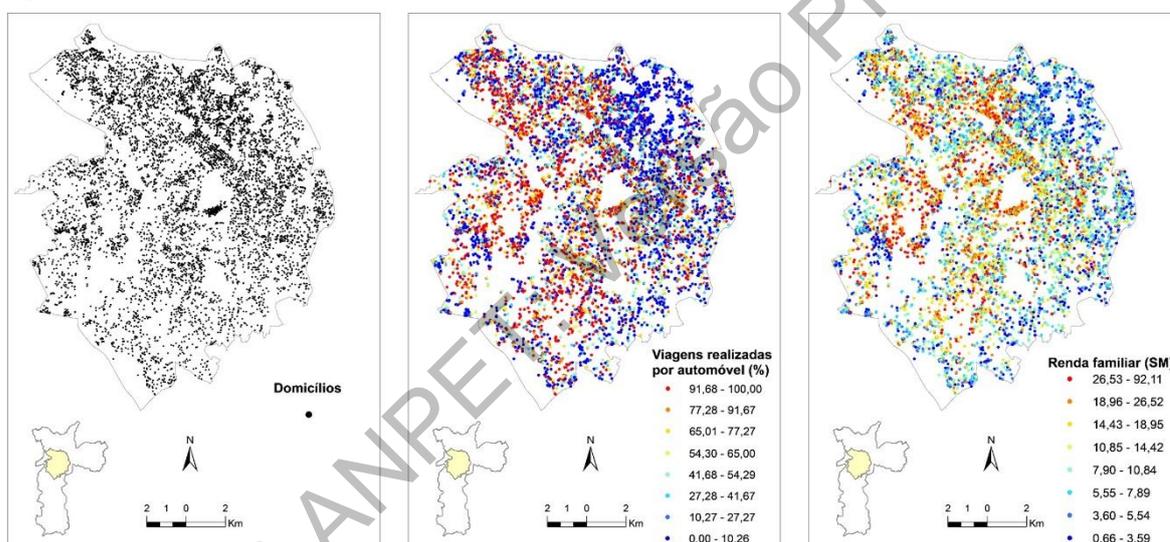
$$[\lambda] = [K]^{-1} \cdot [M] \quad (8)$$

4. ÁREA DE ESTUDO E DADOS

Os dados utilizados neste trabalho são provenientes da Pesquisa Origem-Destino (OD) da Região Metropolitana de São Paulo, realizada em 2007 pela Companhia do Metropolitano de São Paulo – Metrô/SP (Metrô, 2007). Esta pesquisa foi realizada em 469 zonas de tráfego, com uma amostra de, aproximadamente, 30.000 domicílios.

Nesta pesquisa, área de estudo selecionada foi a região central do município de São Paulo. As variáveis utilizadas para o estudo de dados relativos à demanda por transportes foram: *percentual de viagens domiciliares realizadas por automóvel* (a razão entre o número de viagens por automóvel e o total de viagens realizadas por domicílio) e *renda média familiar* (em salários mínimos) (Figura 2b e 2c).

Nos casos onde houve a ocorrência de dois ou mais domicílios no mesmo par de coordenadas geográficas, foi obtida a média aritmética dos valores das variáveis analisadas. Assim, foi considerada apenas uma realização de $z(x)$ em uma dada localização espacial. Após essa etapa, o banco de dados da área analisada passou a ser composto por 8.294 observações (domicílios) (Figura 2a).



(a) Localização dos domicílios

(b) Viagens por automóvel (%)

(c) Renda familiar (SM)

Figura 2: Localização dos domicílios e distribuição espacial das variáveis estudadas. Fonte: Metrô, 2007

5. ETAPAS DO MÉTODO

A Figura 3 sintetiza as etapas metodológicas sequenciais utilizadas neste trabalho. Após a escolha da dimensão das grades para cada variável, os dados pontuais foram agregados através de um *join* espacial, a fim de obter valores médios por quadrícula da respectiva grade (Figura 4). Após esta etapa, foi realizada a aplicação da geoestatística, como a cálculo e ajuste de semivariogramas experimentais e teóricos, respectivamente. Os resultados da validação cruzada e os mapas com valores estimados foram utilizados para comparar o desempenho das estimativas obtidas por meio de dados desagregados e agregados a partir da grade escolhida.

As etapas relativas ao tratamento de dados foram descritas na Seção 4. Os aplicativos utilizados, neste trabalho, encontram-se entre parênteses em algumas etapas, ilustradas na Figura 3.

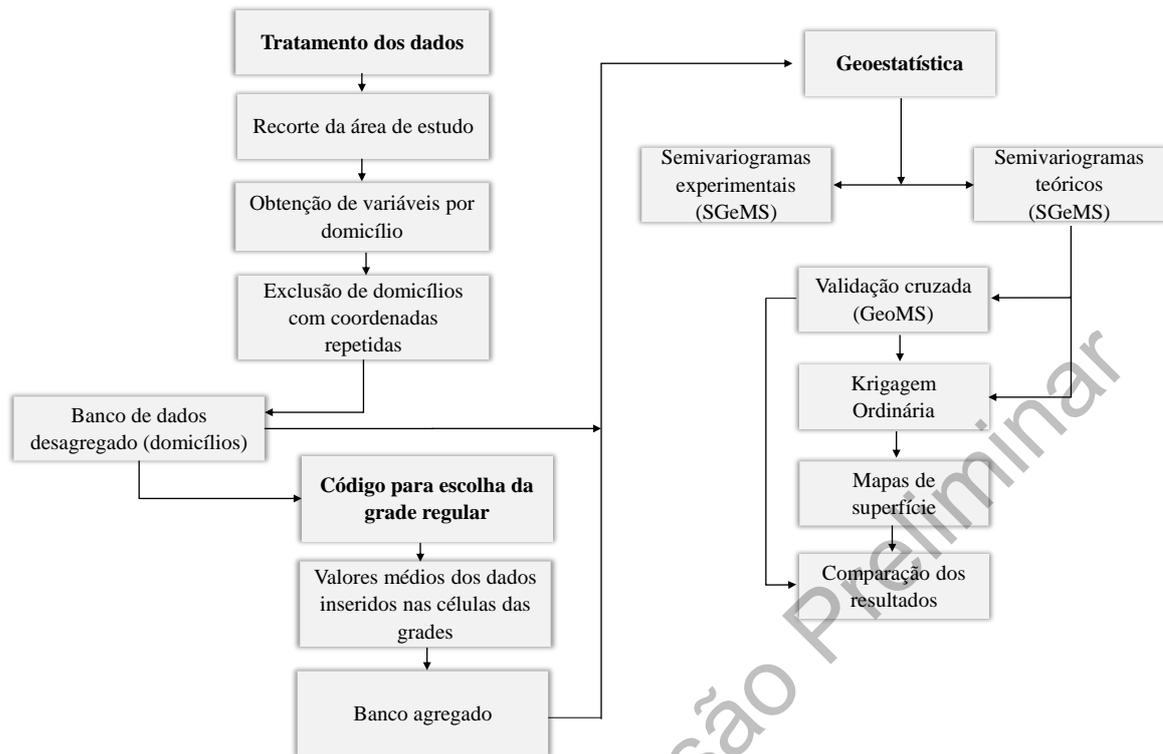


Figura 3: Fluxograma do método

5.1 Escolha da dimensão das grades

O teste não-paramétrico de *Mann–Whitney* compara a distribuição de frequência acumulada entre duas amostras independentes e verifica se elas são similares, sem a necessidade do conhecimento prévio da distribuição populacional de ambas as amostras. O programa *RStudio* (através da função *wilcox.test*) foi utilizado para a elaboração de um código para verificar qual o tamanho de grade (para agregar os dados domiciliares) representaria melhor o comportamento do domicílio. O código gerou, através de *loops*, grades de cinco em cinco metros e obteve valores médios dos pontos inseridos em cada quadrícula.

Para cada configuração de grade gerada foi verificado se a distribuição acumulada dos dois pares de amostra (pontual x grade) eram similares. Para um nível de confiança de 95%, a hipótese nula não foi rejeitada quando o valor $p > 0,05$ (Dinneen e Blakesley, 1973; Hollander e Wolfe, 1999).

Nesse trabalho, a **hipótese nula** afirma que há similaridade de distribuição populacional entre os valores desagregados e os valores médios relativos a cada grade. Por outro lado, a **hipótese alternativa** indica que as distribuições populacionais entre os pares de dados são diferentes. Assim, para cada valor de dimensão de grade foi “tomada a decisão” se a hipótese nula seria rejeitada ou não (Quadro 1). Vale ressaltar que a escolha de uma grade de maior área provê menores erros de estimativa na krigagem. Desta forma, o processo de escolha da melhor escala de representação deve partir da suposição da escolha da maior quadrícula possível, uma vez que possui distribuição populacional similar aos dados originais.

Quadro 1: Resultados do teste não paramétrico de *Mann – Whitney*.

Hipótese nula	<i>Viagens por automóvel (%)</i>			<i>Renda familiar (SM)</i>		
	Valor da grade (m)	p-valor	Decisão	Valor da grade (m)	p-valor	Decisão
"A distribuição é a mesma para as duas amostras"	5	0,91	não rejeita H_0	5	0,91	não rejeita H_0
	10	0,85	não rejeita H_0	10	0,92	não rejeita H_0
	15	0,72	não rejeita H_0	15	0,82	não rejeita H_0
	20	0,64	não rejeita H_0	20	0,72	não rejeita H_0
	25	0,52	não rejeita H_0	25	0,36	não rejeita H_0
	30	0,40	não rejeita H_0	30	0,57	não rejeita H_0
	35	0,25	não rejeita H_0	35	0,60	não rejeita H_0
	40	0,34	não rejeita H_0	40	0,25	não rejeita H_0
	45	0,27	não rejeita H_0	45	0,17	não rejeita H_0
	50	0,16	não rejeita H_0	50	0,25	não rejeita H_0
	55	0,17	não rejeita H_0	55	0,11	não rejeita H_0
	60	0,16	não rejeita H_0	60	0,10	não rejeita H_0
	65	0,08	não rejeita H_0	65	0,07	não rejeita H_0
	70	0,05	não rejeita H_0	70	0,09	não rejeita H_0
	75	0,07	não rejeita H_0	75	0,09	não rejeita H_0
	80	0,05	rejeita H_0	80	0,02	rejeita H_0
	85	0,07	não rejeita H_0	85	0,01	rejeita H_0
	90	0,01	rejeita H_0	90	0,03	rejeita H_0
	95	0,03	rejeita H_0	95	0,01	rejeita H_0
	100	0,03	rejeita H_0	100	0,00	rejeita H_0

5.2 Agregação dos bancos de dados

O objetivo deste trabalho é a determinação do limite de dimensão para uma grade regular, a fim de representar o comportamento individual (desagregado) de variáveis relativas à demanda por transportes. Desta maneira, a grade selecionada foi aquela de maior tamanho e com distribuição populacional similar à pontual, partindo-se do seguinte pressuposto: quanto maior o nível de agregação dos dados, menor será o erro das estimativas obtidas pela krigagem. Como demonstrado no Quadro 1, foram selecionadas as grades de 85 e 75 metros para as variáveis *Viagens por automóvel* e *Renda familiar*, respectivamente. A Figura 4 exemplifica o procedimento de agregação através da obtenção de valores médios por célula.

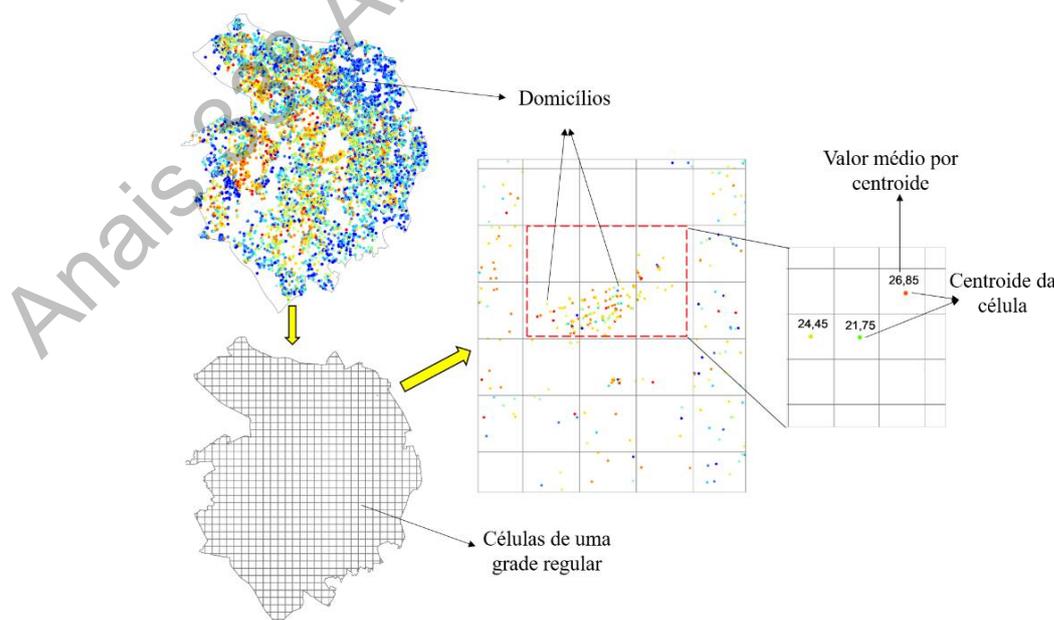


Figura 4: Exemplo da agregação de dados em células de uma grade regular

Apesar da grade de 75 metros ter sido aceita pelo teste de hipótese para a variável referente à escolha modal, utilizou-se a grade de 85 metros, devido à necessidade de escolha da maior grade possível para agregação de dados, a fim de minimizar os erros das estimativas.

6. APLICAÇÃO DA GEOESTATÍSTICA A VARIÁVEIS OBTIDAS POR DIFERENTES ESCALAS

A Figura 5 apresenta os semivariogramas experimentais e as curvas teóricas (referentes para dados desagregados e agregados). O modelo teórico utilizado foi o exponencial, definido através do método dos mínimos quadrados ponderados. Este método de ajuste busca minimizar a diferença média quadrática entre o semivariograma experimental e a curva do modelo teórico.

É possível verificar que os semivariogramas experimentais, obtidos para dados relacionados às grades regulares, apresentam melhor estrutura espacial, em relação aos semivariogramas obtidos através de amostras desagregadas. Observa-se, ainda, considerável redução do efeito pepita, especialmente para o modelo que representa a variável Renda familiar (decréscimo equivalente a 36%). O semivariograma atribuído à variável Percentual de viagens por automóvel (Figura 5a e 5b) apresentou redução menos expressiva deste parâmetro (18%).

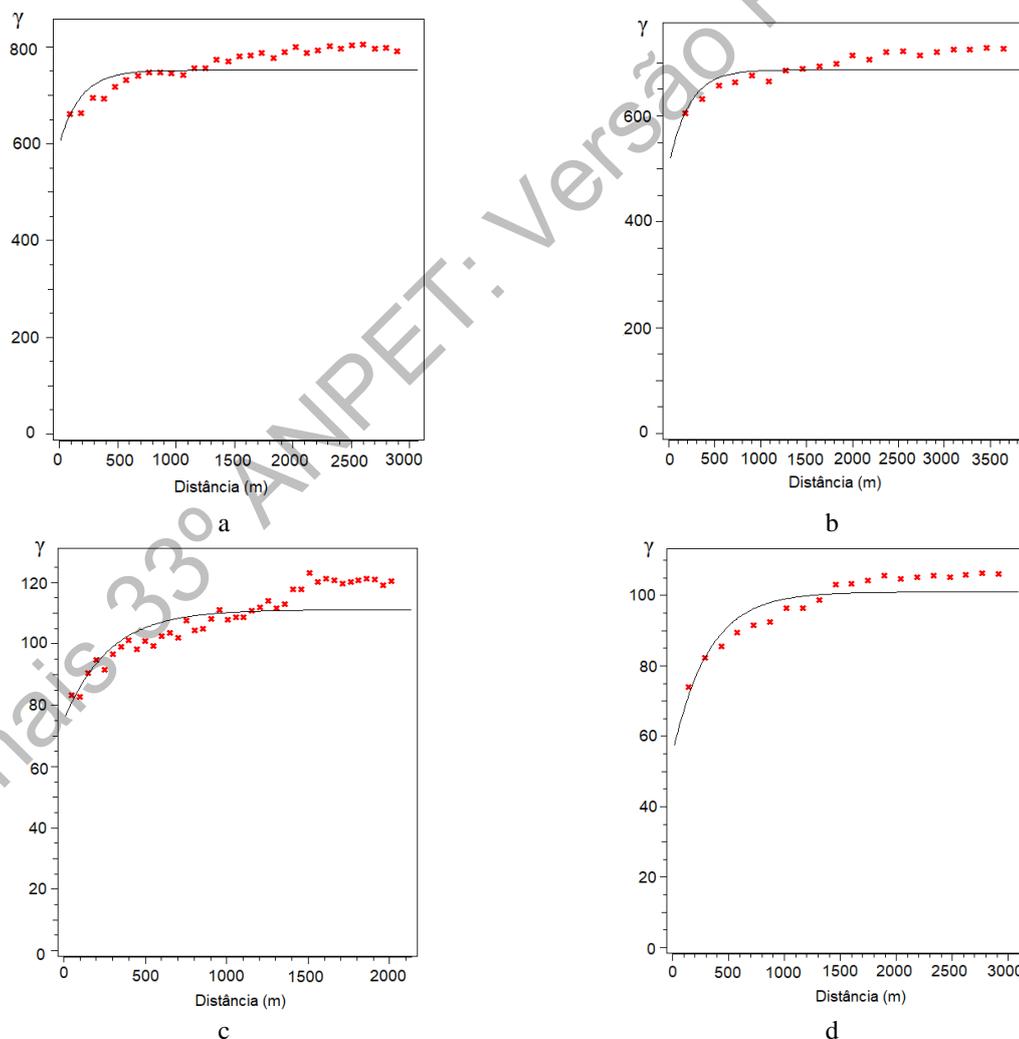


Figura 5: Semivariogramas a) *Percentual de viagens por automóvel – domicílio*; b) *Percentual de viagens por automóvel – grade regular*; c) *Renda familiar*; d) *Renda familiar – grade regular*

6.1 Comparação entre modelos obtidos por diferentes escalas

Após a etapa de análise espacial através dos semivariogramas, foi realizada a etapa de validação cruzada para as variáveis analisadas (desagregadas e agregadas por grades). A Tabela 1 apresenta um resumo das medidas de desempenho para cada um dos modelos obtidos. Vale ressaltar que tais medidas foram calculadas com base nos valores estimados e observados para os dois suportes geográficos.

Tabela 1: Medidas de desempenho dos modelos geoestatísticos

Suporte	Variável	Erro Médio	EQM	Variância média
Domicílios	Auto	23,17	750,97	729,86
	Renda	6,89	93,25	94,18
Grade regular	Auto	21,90	683,27	650,40
	Renda	6,53	83,17	76,85

Se comparados aos ajustes obtidos por dados desagregados, os modelos que representam variáveis agregadas possuem melhor desempenho em relação ao Erro Médio, Erro Quadrático Médio (EQM) e variância média da krigagem.

Os mapas de valores estimados através da krigagem, referentes aos ajustes dos modelos obtidos por dados desagregados e agregados por grade regular, são apresentados na Figura 6. A partir da análise dos dois mapas (modelo desagregado x modelo agregado), nota-se que os mesmos possuem configuração espacial semelhante. Essa constatação é decorrente do nível de agregação proposto, o que gerou dois bancos de dados com comportamento espacial bastante semelhante. Observa-se que os valores mais elevados da variável *Percentual de viagens por automóvel* podem ser comparados, visualmente, aos valores mais elevados da *Renda média (SM)* na área central do município de São Paulo. Esta constatação mostra a considerável influência da renda na escolha do modo de transporte individual motorizado (automóvel), em detrimento dos modos não motorizado e público na área analisada.

7. CONSIDERAÇÕES FINAIS

Este trabalho discutiu a importância do suporte ou escala em análise espacial, além de buscar entender o comportamento individual, utilizando um método para agregar dados pontuais (domicílios ou indivíduos). O objetivo da aplicação do método proposto foi evitar problemas relacionados à mudança de escala, a fim de minimizar possíveis dificuldades no cálculo e modelagem teórica de semivariogramas, ocasionadas pela alta variabilidade de dados representados de forma desagregada.

Pode-se considerar que a definição de uma malha regular influenciou positivamente a qualidade do ajuste dos semivariogramas. A partir da análise geoestatística para dados agregados, através das grades escolhidas, observou-se que houve um relativo declínio da aleatoriedade espacial para as duas variáveis analisadas. Os semivariogramas das variáveis estudadas, *Percentual de viagens por automóvel* e *Renda familiar*, apresentaram uma redução do efeito pepita em 18% e 36%, respectivamente. O alto valor desse parâmetro, associado ao semivariograma teórico, significa que o conjunto amostral possui alta aleatoriedade espacial, dificultando, deste modo, a aplicação da geoestatística. Apesar de ter sido observado o decréscimo deste parâmetro, neste trabalho, a partir da mudança de escala, pode-se considerar que os semivariogramas ainda continuaram apresentando efeito pepita com valores elevados, especialmente para a variável relativa à escolha do automóvel. Essa constatação relaciona-se ao fato de que a agregação dos dados foi realizada por meio de quadrículas pouco espaçadas entre si, resultando em dois bancos de dados (agregado e desagregado) com características muito semelhantes.

A partir do método de validação cruzada, conclui-se que os modelos geoestatísticos das variáveis agregadas por grade regular apresentaram melhor desempenho se comparados aos modelos provenientes de dados desagregados.

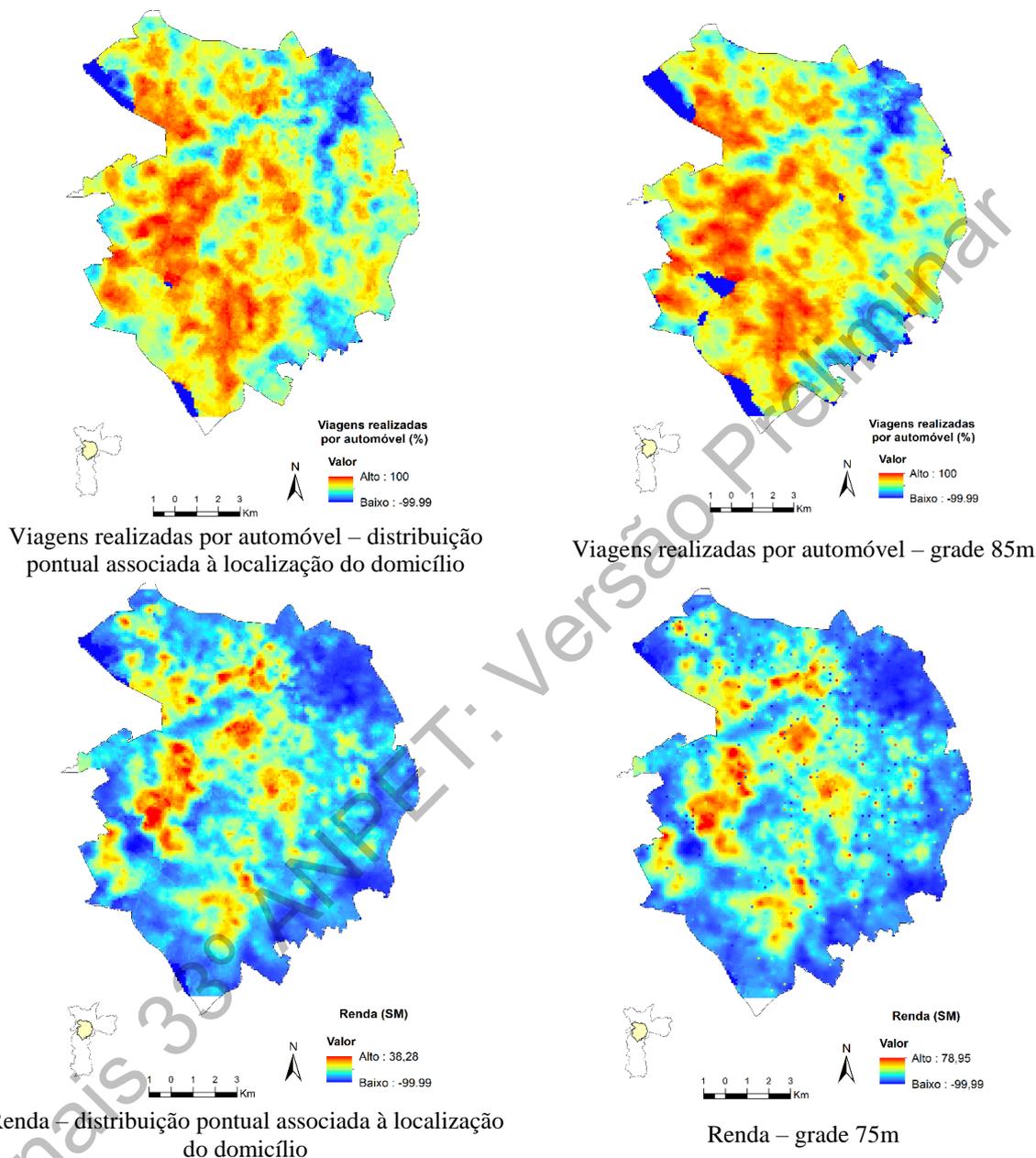


Figura 6: Mapas estimados através da krigagem para a representação desagregada e a grade sugerida.

Agradecimentos

O presente trabalho foi realizado com apoio do CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brasil (Processos 303645/2015-6 e 152536/2018-3).

REFERÊNCIAS BIBLIOGRÁFICAS

- Castro, I. E.; Corrêa, R. L.; Gomes, P. C. G. (1995) *Geografia, conceitos e temas*. Rio de Janeiro: Bertrand Brasil.
- Chilès, J. P.; Delfiner, P. (2012) *Geostatistics Modeling Spatial Uncertainty* (2ª ed.). New Jersey: John Wiley & Sons.
- Cressie, N. (1985) Fitting variogram models by weighted least squares. *Journal of the International Association for Mathematical Geology*, v. 17, n. 5, p. 563–586.

- Dinneen, L. C.; Blakesley, B. C. (1973) Algorithm AS 62: a generator for the sampling distribution of the Mann-Whitney U statistic. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, v. 22, n. 2, p. 269–27.
- Druck, S.; Carvalho, M.S.; Câmara, G.; Monteiro, A. V. M. (2004) *Análise espacial de dados geográficos*. Brasília: Empraba Cerrados.
- Eom, J. K.; Park, M. S.; Heo, T.-Y.; Huntsinger, L. F. (2006) Improving the prediction of annual average daily traffic for nonfreeway facilities by applying a spatial statistical method. *Transportation Research Record: Journal of the Transportation Research Board*, v. 1, n. 1968, p. 20–29, doi: 10.3141/1968-03.
- Fotheringham, S. (1989) Scale-independent spatial analysis. *Accuracy of spatial databases*, p. 221-228, Londres.
- Gelfand, A. E.; Zhu, L.; Carlin, B. P. (2001) On the change of support problem for spatio-temporal data. *Biostatistics*, v. 2, n. 1, p. 31–45.
- Goovaerts, P. (2008) Kriging and Semivariogram Deconvolution in the Presence of Irregular Geographical Units. *Mathematical Geology*, v. 40, n. 1, p. 101–128.
- Gotway, C. A.; Young, L. J. (2002) Combining Incompatible Spatial Data. *Journal of the American Statistical Association*, v. 97, n. 458, p. 632–648.
- Hollander, M.; Wolfe, D. A. (1999) *Nonparametric statistical methods* (2ª ed.). New York: Wiley.
- Holt, D.; Steel, D. G.; Tranmer, M.; Wrigley, N. (1996) Aggregation and ecological effects in geographically based data. *Geographical analysis*, v. 28, n. 3, p. 244–261.
- Isaaks, E. H.; Srivastava, M. R. (1989) *Applied geostatistics*. Oxford University Press.
- Journal, A. G. (1986) Geostatistics: Models and tools for the earth sciences. *Mathematical Geology*, v. 18, n. 1, p. 119–140.
- Journal, A. G.; Huijbregts, C. J. (1978) *Mining geostatistics*. Londres: Academic Press London.
- Klatko, T. J., T. U. Saeed, M. Volovski, S. Labi, J. D. Fricker e K. C. Sinha (2017) Addressing the Local-Road VMT Estimation Problem Using Spatial Interpolation Techniques. *Journal of Transportation Engineering, Part A: Systems*, vol. 143, n. 8. doi:10.1061/JTEPBS.0000064.
- Landim, P. M. B. (2006) Sobre Geoestatística e mapas. *Terra e Didática*, v. 2, n. 1, p. 19–33.
- Lindner, A. (2015) *Análise desagregada de dados de demanda por transportes através de modelagem geoestatística e tradicional*. 106p. Dissertação de mestrado. Departamento de Engenharia de Transportes, Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.
- Matheron, G. (1970) La théorie des variables régionalisées et ses applications. *Les Cahiers du Morphologie Mathématique de Fontainebleau*, v. 5, p. 1–212.
- Metrô. (2007) Companhia do Metropolitano de São Paulo. *Pesquisa Origem e Destino 2007*. Disponível em: <<http://www.metro.sp.gov.br/>>. Acesso em: 10 ago. 2016.
- Marques, S. F.; Pitombo, C. S. (2018) Estimativa do Volume de Passageiros ao Longo de Linhas de Transporte Público por Ônibus a partir da Geoestatística. *Anais do XXXII Congresso de Pesquisa e Ensino em Transportes*, ANPET, Gramado, v.1, p. 2487- 2498.
- Marques, S. F. (2019) *Estimativa do volume de passageiros ao longo de uma linha de transporte público por ônibus a partir da Geoestatística*. Dissertação de mestrado. Departamento de Engenharia de Transportes, Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.
- Openshaw, S. (1976) An Empirical Study of Some Spatial Interaction Models. *Environment and Planning A: Economy and Space*, v. 8, n. 1, p. 23–41.
- Openshaw, S. (1984) The modifiable areal unit problem. *Concepts and techniques in modern geography*.
- Páez, A.; Scott, D. M. (2005) Spatial statistics for urban analysis: a review of techniques with examples. *GeoJournal*, v. 61, n. 1, p. 53–67.
- Pitombo, C. S.; Sousa, A. J.; Birkin, M.; Quintanilha, J. A. (2010) Comparing different spatial data analysis to forecast trip generation. *Anais 12th World Conference on Transport Research Society*. Lisboa, v.1, p. 1-23.
- Pitombo, C. S.; Salgueiro, A. R.; Costa, A. S. G.; Isler, C. A. (2015) A two-step method for mode choice estimation with socioeconomic and spatial information. *Spatial Statistics*, v. 11, p. 45–64.
- Rocha, S. S. (2019) *Uso de Algoritmos Genéticos para otimização de modelagem geoestatística aplicada à demanda por transportes*. Tese de Doutorado. Departamento de Engenharia de Transportes, Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.
- Soares, A. (2000) *Geoestatística para as ciências da terra e do ambiente* (1ª ed.). Lisboa: Instituto Superior Técnico.
- Wackernagel, H. (2003) *Multivariate geostatistics: an introduction with applications* (3ª ed.). Fontainebleau: Springer Science & Business Media.
- Yamamoto, J. K. (2001) *Avaliação e Classificação de Reservas Minerais*. São Paulo: Edusp.
- Yamamoto, J. K.; Landim, P. M. B. (2013) *Geoestatística: conceitos e aplicações*. São Paulo: Oficina de Textos, 2013.