

Estimativa do volume de passageiros ao longo de uma linha de transporte público por ônibus a partir da Geoestatística

Samuel de França Marques¹, Cira Souza Pitombo²

¹Departamento de Engenharia de Transportes, USP, samuelmarques@usp.br

²Departamento de Engenharia de Transportes, USP, cirapitombo@usp.br

Recebido:

10 de abril de 2019

Aceito para publicação:

27 de agosto de 2019

Publicado:

12 de novembro de 2019

Editor de área:

Bruno Vieira Bertoncini

Palavras-chaves:

Estatística espacial,
Pesquisa sobe/desce,
Krigagem,
Demanda por transporte público.

Keywords:

Spatial statistics,
Boarding/Alighting survey,
Kriging,
Transit demand.

DOI:10.14295/transportes.v27i3.2007



RESUMO

A modelagem clássica da demanda por transportes ignora um importante aspecto normalmente presente na estrutura das variáveis de interesse: a autocorrelação espacial. Pesquisas recentes reconhecem e incluem tal característica à estimativa da demanda, mas há limitações referentes aos elementos básicos de tratamento utilizados nas abordagens. No intuito de superar alguns problemas e restrições associados aos estudos anteriores, o presente trabalho se valeu da dependência espacial entre as observações de Embarques e Desembarques, por ponto de parada, e Carregamento nos trechos, ao longo de uma linha de transporte público, a fim de gerar estimativas dessas variáveis em pontos e trechos que não seriam amostrados por ocasião da pesquisa sobe/desce. Tal previsão foi realizada por meio da aplicação de Krigagem Ordinária a uma linha de ônibus da cidade de São Paulo, Brasil. Os resultados obtidos confirmaram a adequabilidade do ferramental geoestatístico à estimativa de variáveis de demanda por transportes ao longo de uma linha de ônibus.

ABSTRACT

The classical travel demand modeling overlooks an important aspect normally found in the variables of interest: spatial autocorrelation. Recent researches recognize and include this characteristic in travel demand forecasting, but there are limitations regarding the basic elements of treatment used in the approaches. In order to overcome some of the problems and constraints associated with previous researches, the present study relied on spatial dependence between Boarding and Alighting observations, per bus stop, and Loading on sections along a public transport line to generate estimates of these variables in stops and sections that would not be sampled during the passenger Boarding and Alighting survey. This prediction was made by applying Ordinary Kriging to a bus line in the city of São Paulo, Brazil. The results confirmed the feasibility of applying Geostatistics to the estimation of travel demand variables along a bus line.

1. INTRODUÇÃO E BACKGROUND

O equilíbrio oferta e demanda por transportes desempenha um papel primordial na economia e qualidade de vida da sociedade. No que tange à demanda por Transporte Público (TP), tal relevância torna-se ainda mais evidente, dado o caráter social intrínseco a esse serviço. Contudo, o balanço entre o volume de usuários e a capacidade do sistema só pode ser atingido a partir de um planejamento essencialmente pautado no diagnóstico da rede de TP em consideração. Essa prática, por sua vez, depende fortemente da realização de pesquisas e levantamentos nas linhas de ônibus coletivo, pontos de parada, estações, domicílios, entre outros elementos envolvidos no sistema. Dentre as pesquisas existentes para prover subsídio ao planejamento da rede de TP, a pesquisa de Embarque e Desembarque (também conhecida como sobe/desce) é a única que consegue levantar informações detalhadas sobre o volume de passageiros ao longo

de uma linha de ônibus. Os resultados provenientes de tal levantamento permitem tomadas de decisão acerca da programação das viagens, aumento ou redução de itinerários, adição ou retirada de pontos de parada etc. (Vuchic, 2005). Além disso, a metodologia de dimensionamento de frota e, conseqüentemente, elaboração do cronograma de viagens, descrita por Ferraz e Torres (2004) e Ceder (2007), exige o conhecimento do volume de passageiros na seção de maior carregamento, dado que só pode ser obtido, com precisão, a partir de uma pesquisa sobre/desce.

Entretanto, levantamentos massivos, como a pesquisa sobre/desce, demandam tempo, planejamento e investimentos financeiros robustos, o que inclui a contratação de uma equipe de profissionais qualificados para a visitação das linhas de ônibus. Por outro lado, em se tratando das variáveis de demanda por transporte público, mais especificamente aquelas oriundas de uma pesquisa de Embarque e Desembarque, nota-se a ocorrência de uma importante característica que as tornam apropriadas para a aplicação da Geoestatística, ferramental que lida justamente com o problema de escassez de dados e estimativa em locais não amostrados. Tal atributo se refere à dependência espacial entre as observações de viagens, também conhecida como autocorrelação espacial, ou seja, ao fato de que a demanda em pontos próximos no espaço tende a apresentar valores semelhantes. Tendo em vista os empecilhos à realização da pesquisa sobre/desce, as técnicas de interpolação espacial empreendidas pela Geoestatística permitiriam que, a partir da demanda referente apenas a determinados pontos de parada e trechos de linhas de ônibus visitados por ocasião da pesquisa de Embarque e Desembarque, fosse possível encontrar os valores dessa variável nos pontos e trechos não amostrados. Essa solução, portanto, eliminaria a necessidade de realização do referido levantamento em todo o percurso das linhas de ônibus e/ou todos os itinerários da rede, mas ainda garantiria resultados aproximados aos da pesquisa completa.

Apesar de ter sido criada, inicialmente, para tratar de variáveis relacionadas à mineração (Krige, 1951), os benefícios fornecidos pela Geoestatística fizeram com que seus conceitos fossem expandidos a diversas áreas de estudo (Bayraktar e Turalioglu, 2005; Goovaerts, 2012; Oliver e Webster, 1986; Zimmerman et al., 1998). Na engenharia de transportes, o emprego dos conceitos geoestatísticos já foi verificado na modelagem de acidentes/segurança viária (Gomes et al., 2018; Gundogdu, 2014; Majumdar et al., 2004; Manepalli e Bham, 2011) e, mais recentemente, na estimativa de variáveis de demanda por transportes. Nesse sentido, foram desenvolvidos trabalhos que abordam viagens no âmbito de zonas de tráfego (Lindner et al., 2016; Rocha et al., 2017) e domicílios (Gomes et al., 2016; Lindner e Pitombo, 2017; Pitombo et al., 2015), que, devido ao nível de agregação presente nas variáveis de interesse, não conseguiram alcançar resultados totalmente satisfatórios. No caso da demanda agregada por zonas de tráfego, o pressuposto da Geoestatística de que tanto os elementos a serem utilizados para as estimativas quanto aqueles onde se pretende prever determinada variável possuem o mesmo tamanho e forma (Goovaerts, 2008) é violado. Por outro lado, quando as observações de viagens são dadas por indivíduo ou domicílio, a alta aleatoriedade atrelada ao comportamento humano interfere sobremaneira no desempenho das técnicas.

Em suma, porém, a maioria dos trabalhos referentes à demanda por transportes, para dados em rede, se concentra na aplicação da Geoestatística à estimativa/modelagem do Volume Médio Diário Anual (VMDA) em trechos viários que não possuem postos de contagem de tráfego (Eom et al., 2006; Klatko et al., 2017; Selby e Kockelman, 2013; Shamo et al., 2015; Subhadip et al., 2017; Wang e Kockelman, 2009). Contudo, mesmo considerando que, nesse caso, a unidade básica de agregação (trechos viários) é muito mais regular que no caso das zonas de tráfego e

menos desagregada em comparação aos indivíduos e domicílios, os resultados não foram suficientemente bons. Isso se deve, entre outros motivos, ao fato de que as informações utilizadas para a modelagem geoestatística estão, no que concerne a trechos viários, distribuídas irregularmente no interior do espaço a ser interpolado. Dessa forma, apenas regiões com uma alta densidade de postos, normalmente aquelas situadas próximas a áreas urbanas, conseguem gerar números precisos. Em contrapartida, no que tange a redes de transporte público por ônibus, os pontos de parada, que exercem a mesma função que os postos de contagem de tráfego no caso do VMDA, estão densamente distribuídos no território urbano.

No que tange à aplicação da Geoestatística à demanda por transporte público urbano, apenas o estudo de Zhang e Wang (2014) foi encontrado até o presente momento. Esse trabalho, no entanto, além de se referir ao transporte público por metrô, cujas estações estão separadas por uma distância superior ao espaçamento médio entre pontos de parada, lidou apenas com a variável de Embarques. Tendo em vista que, nos sistemas de metrô, é possível acompanhar automaticamente a quantidade de usuários que entram nas estações, os valores coletados para o número de Embarques foram utilizados somente para prever a movimentação de passageiros em uma nova linha a ser implantada. No caso do transporte público por ônibus, entretanto, tal variável pode ser levantada, com precisão, apenas por meio da pesquisa sobre/desce. Os contadores automáticos de passageiros (Barbosa et al., 2006) são uma alternativa promissora, mas, devido ao seu elevado custo, ainda não foram popularizados. Há, também, soluções estratégicas em que os dados da bilhetagem eletrônica são associados aos de localização dos veículos de transporte público (GPS), para os casos em que tais tecnologias estão disponíveis. Contudo, o método para a estimativa dos pontos de Embarque e Desembarque de cada passageiro exigiria a adoção de suposições que comprometem a precisão dos resultados, sendo que, nesse caso, levantamentos complementares ainda seriam necessários (Bagchi e White, 2005).

Finalmente, levando-se em conta os trabalhos mencionados anteriormente, é possível enumerar as seguintes lacunas de pesquisa, alinhadas com os objetivos do presente trabalho: (1) estimar variáveis de demanda por transportes, a partir da Geoestatística, com suportes homogêneos (pontos de parada, por exemplo) e baixa variabilidade espacial associada a tais suportes; (2) estimar variáveis pertencentes a uma rede de transporte urbana e não intermunicipal (como VDMA – trabalhos anteriores); e (3) avaliar o desempenho das técnicas de estatística espacial em estimar/prever o valor de variáveis de demanda por transporte público (número de Embarques e Desembarques por ponto de parada e ocupação nos trechos), supondo que, para uma determinada rede de TP, nem todos os pontos, trechos e/ou linhas de ônibus foram amostrados.

Dessa forma, o objetivo principal desse trabalho reside em avaliar o desempenho da Krigagem Ordinária (KO), técnica de interpolação geoestatística que exige apenas o valor da variável de interesse em pontos amostrados no espaço e suas respectivas coordenadas geográficas, na estimativa das três variáveis de demanda catalogadas em uma pesquisa sobre/desce tradicional, a saber: número de Embarques e de Desembarques por ponto de parada e Carregamento nos trechos delimitados por dois pontos de parada consecutivos. Tal análise será realizada por meio de um estudo de caso real, dirigido a uma linha do sistema de transporte público por ônibus da cidade de São Paulo-SP.

Este artigo possui 4 seções, além dessa introdução. A segunda seção introduz o banco de dados utilizado no estudo de caso e o procedimento metodológico a ele aplicado. Posteriormente, os resultados correspondentes a cada etapa do método são apresentados e discutidos,

em sequência, na Seção 3. As contribuições alcançadas pela solução proposta ao problema abordado são destacadas na Seção 4, a qual também traz sugestões para trabalhos futuros, a fim de fortalecer, consolidar e aprofundar as conclusões encontradas nesse trabalho.

2. MATERIAIS E MÉTODO

A seguir, introduzem-se, em sequência, o banco de dados coletado para a realização do estudo de caso, os tratamentos aplicados a ele, as ferramentas computacionais utilizadas e, finalmente, as etapas que compõem a análise e solução do problema proposto.

2.1. Base de dados e recursos computacionais

Para verificar o desempenho das estimativas geoestatísticas na previsão da demanda por transporte público, se faz necessária a aquisição de um banco de dados, proveniente de uma pesquisa sobre/desce, no qual estejam identificados os pontos de parada e suas respectivas coordenadas geográficas. A busca por uma cidade que dispusesse de tais informações culminou na adoção da capital paulista para a realização do estudo de caso.

De acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE, 2018), São Paulo é o município brasileiro com maior população, consolidando um total de 12.106.920 habitantes em 2017 e densidade demográfica de 7.398,26 hab/km². Além disso, também é a cidade com maior Produto Interno Bruto, cujo valor provém, principalmente, do setor de serviços e industrial, com participação, também, da administração pública e agropecuária. No que tange à forma como a população se locomove no município, a Pesquisa de Mobilidade de 2012 revelou que 38,96% das viagens eram realizadas por modos coletivos, com uma representatividade do ônibus superior à metade desses deslocamentos (São Paulo, 2015).

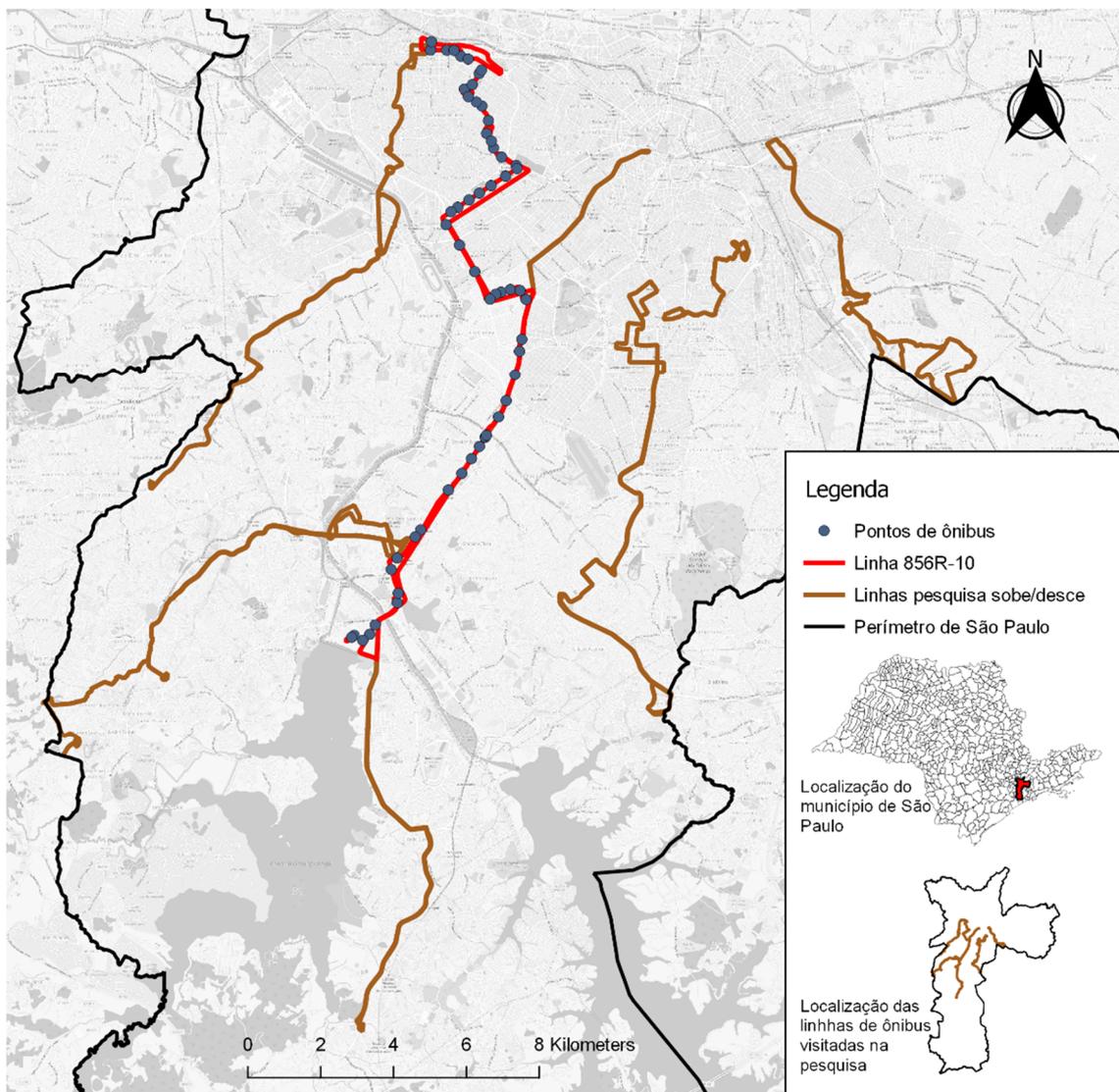
O sistema de transporte público de São Paulo, gerenciado pela São Paulo Transporte S.A. (SPTrans), é composto por um total de 1.355 linhas de ônibus e 20.006 pontos de parada (SPTrans, 2018). Dessa rede, foram disponibilizadas, pela SPTrans, planilhas contendo os resultados de uma pesquisa de Embarque e Desembarque realizada em oito linhas de ônibus da capital paulista em 2012 e uma extrapolação para o ano de 2017 a partir de dados da bilhetagem eletrônica do sistema de transporte público por ônibus da cidade.

Além de incluir o número de Embarques, de Desembarques e Carregamento dos trechos a jusante dos pontos de parada para seis faixas de horários, as planilhas ainda possuíam um código identificador de cada ponto de parada da linha, bem como sua posição na sequência da viagem. Informações acerca das linhas de ônibus também foram fornecidas, tais como: código identificador da linha e da viagem e um número associado ao sentido da viagem, ou seja, ida ou volta. A espacialização desse banco de dados foi realizada, posteriormente, a partir de informações do Sistema de Consulta do Mapa Digital da Cidade de São Paulo, hospedado no sítio eletrônico do GeoSampa, no qual teve-se acesso às malhas digitais (arquivos *shapefiles*) das linhas de ônibus e pontos de parada de São Paulo.

De posse dos códigos identificadores das oito linhas de ônibus contempladas pela pesquisa sobre/desce e de seus respectivos pontos de parada, foi possível filtrar, a partir dos *shapefiles* disponíveis no sítio do GeoSampa, apenas aquelas linhas e pontos que interessariam ao presente trabalho. Após esse procedimento, o *shapefile* correspondente às oito linhas de ônibus, que continha feições geométricas do tipo linha contínua para cada sentido de viagem, foi segmentado no intuito de gerar trechos entre pontos de parada consecutivos. A esses segmentos de linha, atribuiu-se, então, a ocupação de passageiros equivalente, proveniente da planilha da SPTrans. Porém, haja vista que a Geoestatística foi desenvolvida para lidar com elementos na

forma de pontos, fez-se necessária, também, a alocação do Carregamento em cada trecho a um ponto situado ao longo do respectivo segmento de linha. O escolhido para carregar essa informação foi o ponto médio dos ramos de linha, uma vez que o centroide geométrico dos trechos se localizou, em alguns casos, fora da extensão do segmento.

A partir de uma inspeção visual do comportamento espacial das oito linhas com os dados de 2017, a linha 856R-10, em seu sentido de ida, foi escolhida para compor o estudo de caso. Tal seleção se deu por meio da análise dos mapas temáticos de Embarques, Desembarques e Carregamento, que, para a linha supracitada, apresentaram uma estrutura espacial aparente que poderia gerar resultados satisfatórios tão somente por meio da aplicação da Krigagem Ordinária. Outros critérios que também nortearam a decisão pela escolha da linha 856R-10 foram: número razoável de pontos de parada em um sentido (quanto maior o número de pontos, maiores são as chances de se obter resultados precisos) e cobertura espacial da linha (itinerário abrangendo diversas regiões da cidade). A Figura 1 mostra o itinerário das oito linhas de ônibus visitadas por ocasião da pesquisa de Embarque e Desembarque, destacando, em vermelho, a linha 856R-10 e respectivos pontos de parada (no sentido de ida), bem como sua localização no interior do município de São Paulo.



Em suma, consolidaram-se os seguintes dados: coordenadas geográficas e número de Embarques e Desembarques por ponto de parada; volume de passageiros por trecho e coordenadas geográficas do ponto médio de cada segmento de linha; e distância em linha reta e ao longo da linha de ônibus entre cada ponto do banco e os pontos remanescentes. As informações de demanda correspondem somente ao percurso de ida da linha 856R-10 e ao conjunto de viagens realizadas no período das 05:00 às 08:59, que abrange o horário de pico da manhã, em um dia típico (terça-feira) de 2017. A esses dados, portanto, foram aplicados os conceitos apresentados a partir da subseção 2.2.

Vale ressaltar que os procedimentos que compõem o método proposto nesse trabalho foram concebidos através do apoio de várias ferramentas computacionais, tais como: IBM SPSS 24.0 para análise exploratória tradicional (IBM, 2016); R (R Core Team, 2018; Millard, 2013) para transformação de Box-Cox; GeoDa (Anselin, 2004; Anselin, Syabri e Kho, 2005) para análise de dependência espacial; e ArcGIS 10.1 para modelagem geoestatística (ESRI, 2010).

2.2. Análise exploratória

A fim de se conhecer melhor as características do banco de dados em consideração, foi realizada uma análise exploratória das variáveis de interesse a partir do cálculo das medidas descritivas tradicionais de tendência central e dispersão. No que concerne à demanda por transportes, usualmente fornecida na forma de contagens, nota-se, em geral, a ocorrência de uma assimetria positiva bastante proeminente, ou seja, fuga de normalidade. Essa peculiaridade pode comprometer sobremaneira o desempenho de alguns estimadores geoestatísticos, tais como a Krigagem Ordinária, que assume distribuição gaussiana para as variáveis de interesse.

No intuito de contornar esse problema, as três variáveis objeto de estudo do presente trabalho (número de Embarques e Desembarques por ponto de parada e Carregamento dos trechos de linhas de ônibus) foram submetidas à transformação de Box-Cox (Box e Cox, 1964). Conhecidas como transformações da família de potência, a variável resposta y é convertida em Z conforme a Equação 1.

$$Z = y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda \dot{y}^{\lambda-1}} & \lambda \neq 0 \\ \dot{y} \ln y & \lambda = 0 \end{cases} \quad (1)$$

em que $\dot{y} = \ln^{-1}[(1/n)\sum \ln y]$ é a média geométrica das n observações e λ é um parâmetro a ser estimado. Tendo em vista que o objetivo da transformação é aproximar a distribuição dos dados à normalidade, o parâmetro λ pode ser encontrado a partir da otimização de alguma função objetivo relacionada à distribuição de Gauss. A alternativa selecionada para esse fim foi o coeficiente de correlação entre as observações da variável transformada e a reta da distribuição normal no gráfico quantil-quantil (Vogel, 1986). Cabe ressaltar, entretanto, que como para um valor de λ nulo, a transformação de Box-Cox equivale à utilização do logaritmo natural dos dados brutos, foi necessário somar uma unidade às variáveis de Embarques e Desembarques por ponto de parada, a fim de se evitar problemas com a indeterminação proveniente do logaritmo de zero, solução sugerida por Bartlett (1947). Posteriormente, no intuito de confirmar a normalidade das variáveis transformadas, aplicaram-se os testes de Kolmogorov-Smirnov e Shapiro-Wilk (Shapiro e Wilk, 1965).

Após essa operação, as variáveis resultantes da transformação de Box-Cox passaram por uma etapa de análise exploratória espacial, cujo objetivo era obter indícios da potencial adequabilidade das técnicas geoestatísticas ao problema abordado no presente trabalho. Nesse âmbito, o

índice de Moran (Moran, 1948), já bastante consolidado em diversas áreas de estudo, surge como uma alternativa promissora. A fórmula para o cálculo desse indicador é mostrada na Equação 2.

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (Z_i - \bar{Z})(Z_j - \bar{Z})}{\sum_{i=1}^n (Z_i - \bar{Z})^2} \quad (2)$$

em que \bar{Z} média aritmética da variável Z ;

w_{ij} ponderador que reflete a intensidade de interação espacial entre os pontos i e j . Baseia-se na distância entre os pontos do banco de dados.

A exemplo do coeficiente de correlação linear de *Pearson*, o índice de Moran também varia de -1 a 1, sendo que o extremo negativo reflete a máxima dispersão espacial passível de ocorrer em uma Variável Regionalizada. Em outras palavras, todos os pontos contidos no banco de dados exibem um comportamento totalmente oposto ao de seus vizinhos imediatos. A condição nula expressa inexistência de estrutura espacial, enquanto que o valor máximo, 1, corresponde ao perfeito estado de autocorrelação espacial, isto é, aquele em que todos os pontos adjacentes entre si apresentam valores semelhantes. Esse último cenário é o mais apropriado/indicado para aplicações da Geoestatística, conceituada sucintamente nas próximas subseções.

2.3. Cálculo e ajuste do semivariograma empírico

O semivariograma $\gamma(h)$, ou variograma $2\gamma(h)$, é a principal ferramenta gráfica da Geoestatística, pois permite visualizar a estrutura espacial da variável de interesse. A partir da hipótese de estacionariedade de segunda ordem (Matheron, 1971), a variância das observações deixa de depender da localização geográfica dos pontos amostrais e passa a se relacionar apenas à distância entre pares de pontos do banco de dados. Dessa forma, o cálculo do semivariograma empírico, ou experimental, é dado pela Equação 3.

$$\gamma(h) = \frac{1}{2N} \sum_{i=1}^N [Z(x_i + h) - Z(x_i)]^2 \quad (3)$$

em que $Z(x)$ valor da Variável Regionalizada Z na posição geográfica amostrada x ;

N número de pares situados a uma distância h de x .

Para a aplicação da Equação 3 aos pontos amostrados, é necessário, inicialmente, definir cinco parâmetros que interferem nos valores encontrados, a saber: direção (azimute), tolerância angular, passo, tolerância do passo e largura máxima. Tais especificações ajudam a definir uma janela no espaço dentro da qual busca-se encontrar eventuais pontos que compõem, juntamente com os pontos de referência, os pares sobre os quais o semivariograma será calculado (Oliver e Webster, 2015; Yamamoto e Landim, 2015). Em se tratando do presente trabalho, utilizou-se o semivariograma omnidirecional, isto é, aquele que seleciona todos os pares separados de uma distância h , independentemente da direção desse vetor. Apesar da linha 856R-10 se estender por uma orientação quase bem definida e, conseqüentemente, ter uma direção principal bastante aparente, não é possível considerar anisotropia para a krigagem neste caso, considerando a impossibilidade de cálculo do semivariograma experimental para a direção ortogonal por ausência de pontos. Assim, os autores optaram pelo cálculo de semivariogramas omnidirecionais.

Em seguida, ajusta-se um modelo teórico aos valores resultantes do semivariograma experimental. O processo de ajuste de uma função bem definida aos pontos do semivariograma experimental consiste da obtenção de três principais parâmetros, efeito pepita, contribuição e alcance, a partir de determinado método pré-estabelecido. O efeito pepita (C_0) refere-se ao valor da semivariância quando a distância entre dois pontos é zero, ou seja, mede a continuidade e irregularidade do fenômeno regionalizado próximo à origem. Segundo Yamamoto e Landim (2015), a ocorrência de efeitos pepitas diferentes da nulidade pode resultar tanto de uma variância aleatória da variável espacial quanto da escala de amostragem, podendo incluir, também, eventuais erros de medição. Já a contribuição (C), ou patamar parcial, reflete o nível de dependência espacial dos pontos conforme a distância entre eles aumenta. O alcance (a/A), por sua vez, pode ser definido como a distância máxima dentro da qual existe autocorrelação espacial entre os pontos. Os três principais modelos teóricos de semivariograma são mostrados nas equações da Tabela 1.

Tabela 1 - Modelos típicos de semivariograma

Nome	Equação
Exponencial	$\gamma(h) = C_0 + C \left[1 - \exp\left(-\frac{h}{a}\right) \right] \quad (4)$
Gaussiano	$\gamma(h) = C_0 + C \left[1 - \exp\left(-\left(\frac{h}{a}\right)^2\right) \right] \quad (5)$
Esférico	$\gamma(h) = \begin{cases} C_0 + C \left[1,5\frac{h}{a} - 0,5\left(\frac{h}{a}\right)^3 \right] & \text{para } h < a \\ C_0 + C & \text{para } h \geq a \end{cases} \quad (6)$

Fonte: Adaptado de Olea (2006)

Nota: C_0 , C e a equivalem, respectivamente, ao efeito pepita, patamar parcial e alcance.

O ajuste do semivariograma empírico a esses três modelos teóricos foi efetuado por meio dos mínimos quadrados ponderados (WLS, do inglês *Weighted Least Squares*), método introduzido por Cressie (1985). Porém, ao invés do inverso da variância a cada distância h , utilizou-se, como peso, o número de pares incluídos no cálculo de cada ponto do semivariograma empírico. Dessa forma, ponderadores maiores são atribuídos às médias obtidas a partir de um número superior de combinações, o que torna esse procedimento mais adequado ao ajuste do semivariograma teórico. As três curvas teóricas supracitadas, que são as mais comumente usadas, foram escolhidas no intuito de se verificar se algum desses modelos demonstra um ajuste muito melhor ao comportamento das variáveis em comparação com os outros.

2.4. Validação cruzada

No que tange a estimativas geoestatísticas, a validação cruzada é realizada pelo teste do ponto fictício. Essa técnica consiste em retirar pontos do banco de dados um por vez e calcular seu valor a partir dos valores dos pontos remanescentes e dos parâmetros do semivariograma teórico (Cressie, 1993). Logo, de posse do valor real observado nos pontos e de seu respectivo valor estimado, é possível estabelecer diversas métricas de erros a fim de se avaliar o desempenho da ferramenta de estatística espacial aplicada. As equações de algumas das principais medidas de aderência estão apresentadas na Tabela 2.

Tabela 2 - Métricas de aderência

Nome	Equação	Comentários
Erro Quadrado (SE)	$\sum_{i=1}^n [y(s_i) - y_{-i}^*(s_i)]^2 \quad (7)$	-
Erro Médio (ME)	$\frac{1}{n} \sum_{i=1}^n [y_{-i}^*(s_i) - y(s_i)] \quad (8)$	Indica a existência de erros sistemáticos
Erro Médio Absoluto (MAE)	$\frac{1}{n} \sum_{i=1}^n y(s_i) - y_{-i}^*(s_i) \quad (9)$	Não particularmente sensível a grandes erros
Raiz Quadrada do Erro Quadrado Médio (RMSE)	$\sqrt{\frac{1}{n} \sum_{i=1}^n [y(s_i) - y_{-i}^*(s_i)]^2} \quad (10)$	Grandes erros são duramente penalizados
Coefficiente de correlação linear de Pearson (R)	$\frac{1}{n} \sum_{i=1}^n \frac{[y(s_i) - \bar{y}][y_{-i}^*(s_i) - \bar{y}^*]}{\sigma_y \sigma_{y^*}} \quad (11)$	-
Taxa de erro (%)	$\frac{y_{-i}^*(s_i) - y(s_i)}{y(s_i)} * 100 \quad (12)$	-

Fonte: Adaptada de Hollander e Liu (2008)

Nota: $y(s_i)$ e $y_{-i}^*(s_i)$ representam, respectivamente, os valores observados e estimados. O subscrito $-i$ alerta que o valor real da variável no ponto i não é usado no cálculo da estimativa em i .

Em se tratando das variáveis de interesse do presente trabalho, a validação cruzada foi utilizada como ferramenta para a definição dos parâmetros do semivariograma teórico e empírico que resultavam no menor erro quadrado médio, para os três modelos considerados: exponencial, gaussiano e esférico. Além disso, também foram calculadas as métricas consolidadas na Tabela 2, na qual y e y^* são, respectivamente, os valores reais e aqueles obtidos por meio do teste do ponto fictício, ambos já resultantes da transformação inversa. Após definido o melhor semivariograma (exponencial, esférico ou gaussiano) para as três variáveis de interesse, seguiu-se para a etapa de interpolação espacial por meio da Krigagem Ordinária.

2.5. Krigagem

O processo de krigagem de uma Variável Regionalizada consiste basicamente na obtenção de pesos ótimos para a combinação linear de ponderadores e valores vizinhos que origina uma superfície contínua de pontos estimados, abrangendo, também, as localizações não amostradas. Em síntese, o estimador da Krigagem Ordinária é dado pela Equação 13.

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (13)$$

em que $Z^*(x_0)$ valor estimado da Variável Regionalizada Z na posição geográfica x_0 ;

λ_i peso ótimo atribuído pela krigagem à observação do vizinho i .

A busca por um estimador que evitasse erros sistemáticos em razão de amostragens irregulares e/ou deficientes do fenômeno regionalizado fez com que os ponderadores oriundos da krigagem fossem considerados ótimos, isto é, que dessem origem a valores previstos não enviesados e com variância mínima (Matheron, 1971). Tal resultado é alcançado a partir da determinação de que: 1) a esperança matemática da diferença entre os valores observados e estimados é nula; e 2) o somatório dos pesos λ_i deve ser igual a 1 (Yamamoto e Landim, 2015). Dessa forma, além de permitir inferência estatística nas estimativas, recurso ausente em

técnicas de interpolação espacial anteriores à Geoestatística, a krigagem ainda consegue gerar intervalos de confiança mais estreitos.

O formalismo estatístico imbuído no estimador da KO, que abrange a estrutura da Equação 13 e as duas restrições mencionadas anteriormente, estabelece o sistema de equações da Krigagem Ordinária em termos da função semivariograma, conforme mostrado na Equação 14 (Yamamoto e Landim, 2015).

$$\begin{cases} \sum_{j=1}^n \lambda_j \gamma(h_{x_i-x_j}) + \mu = \gamma(h_{x_0-x_i}) & \text{para } i = 1, n \\ \sum_{j=1}^n \lambda_j = 1 \end{cases} \quad (14)$$

Em que μ corresponde à média local da Variável Regionalizada e o termo entre parênteses representa a distância entre as duas localizações x . O sistema de equações 14 resulta da operação matricial detalhada a seguir.

$$\begin{bmatrix} \gamma(h_{1-1}) & \gamma(h_{1-2}) & \dots & \gamma(h_{1-n}) & 1 \\ \gamma(h_{2-1}) & \gamma(h_{2-2}) & \dots & \gamma(h_{2-n}) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(h_{n-1}) & \gamma(h_{n-2}) & \dots & \gamma(h_{n-n}) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma(h_{0-1}) \\ \gamma(h_{0-2}) \\ \vdots \\ \gamma(h_{0-n}) \\ 1 \end{bmatrix}$$

A matriz à esquerda compreende o valor da semivariância teórica entre os n pontos amostrados. Essa medida é calculada substituindo-se a distância entre todos os pares do banco de dados nas equações já calibradas dos modelos exponencial, gaussiano e esférico. Cabe lembrar que a diagonal principal dessa matriz será composta pelo efeito pepita obtido para a respectiva curva. A matriz central, por sua vez, se refere aos pesos ótimos da krigagem e à média local onde pretende-se estimar a variável de interesse. O produto entre a matriz de pesos (Λ) e a de semivariância entre pontos amostrados (Γ) dá origem à matriz de semivariância entre o ponto a ser estimado e seus n vizinhos (W), também obtida a partir do semivariograma teórico. Logo, a resolução do sistema de equações 14 pode ser simplificado em notação matricial conforme exposto na Equação 15 (Cressie, 1993; Matheron, 1971).

$$\Lambda = \Gamma^{-1}W \quad (15)$$

Tendo em vista a simplicidade da Krigagem Ordinária, mas a robustez dos resultados empreendidos por ela, esse foi o estimador geoestatístico escolhido para compor a presente análise.

3. RESULTADOS E DISCUSSÕES

A Figura 2 exibe o perfil de movimentação de passageiros ao longo dos 57 pontos de parada e 56 primeiros trechos da linha 856R-10, no sentido de ida, referente ao conjunto de viagens realizadas no período das 05:00 às 08:59 em um dia típico (terça-feira) de 2017.

A Figura 2 permite visualizar o motivo que dá suporte à escolha da linha 856R-10, em seu percurso de ida, para a realização do estudo de caso. O trajeto se inicia na extremidade esquerda dos mapas e termina na margem direita, cobrindo uma extensão de aproximadamente 26 km. Dessa forma, é perceptível a estrutura espacial presente, principalmente, em Embarques e Carregamento, cujos maiores volumes de passageiros se concentram na primeira metade de pontos e trechos, respectivamente, e vão diminuindo conforme os veículos completam o itinerário. Apesar de Desembarques demonstrar um comportamento semelhante ao de Embarques, nota-se que essa variável não demonstra um padrão tão regular quanto o da primeira, haja vista a

ocorrência de um elevado volume de Desembarques também em pontos intermediários da linha, nos quais o número de Embarques é relativamente pequeno, e em alguns pontos localizados próximos ao fim do itinerário.

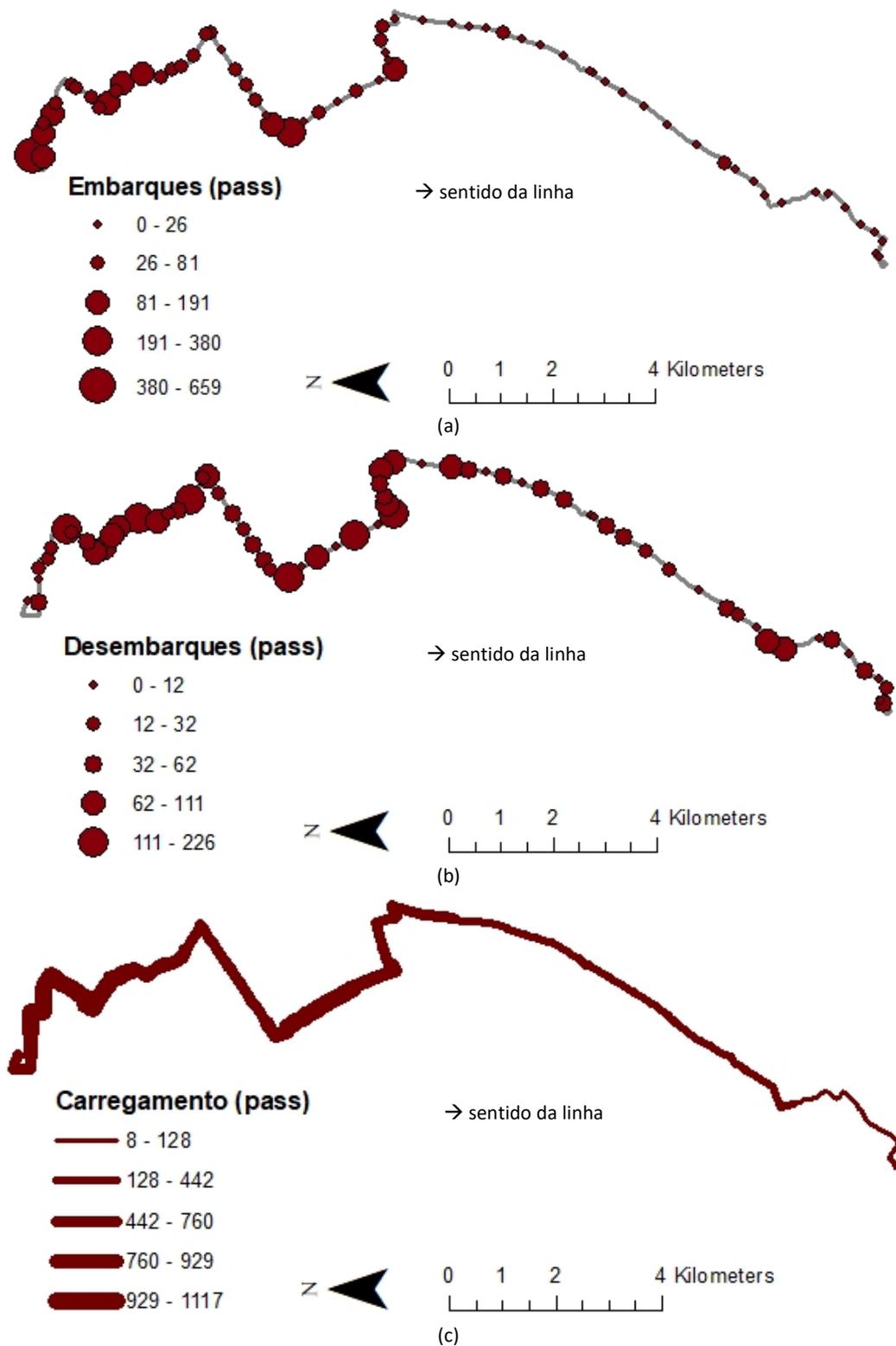


Figura 2. Perfis de a) Embarques, b) Desembarques e c) Carregamento na linha 856R-10

A Tabela 3 consolida as medidas de estatística descritiva tanto para as variáveis de interesse em sua forma bruta quanto transformadas pelo método de Box-Cox. Cabe mencionar que a potência ótima encontrada para Embarques, Desembarques e Carregamento foi de 0,107, 0,250 e 1,152, respectivamente, sendo que o coeficiente de correlação do gráfico de probabilidade das três variáveis aumentou após a transformação.

Tabela 3 - Medidas descritivas das variáveis de interesse

Descritivos\Variável	Embarques	Embarques Box-Cox	Desembarques	Desembarques Box-Cox	Carregamento	Carregamento Box-Cox
n observações	57	57	57	57	56	56
Média	58,81	1,44	58,95	2,59	620,36	1.691,87
Desvio Padrão	101,50	0,20	54,24	0,61	330,83	983,94
Mínimo	0,00	1,00	0,00	1,00	38,00	65,99
25%	13,00	1,33	23,00	2,21	323,50	777,42
50%	28,00	1,43	41,00	2,55	723,00	1.962,94
75%	59,50	1,55	77,50	2,98	918,75	2.586,83
Máximo	659,00	2,00	226,00	3,89	1.117,00	3.239,53

É importante notar que a diferença de uma unidade entre o número de observações das variáveis de Carregamento se comparadas às demais se deve ao fato de que o trecho a jusante do último ponto de parada pertence ao percurso de volta da linha. Cabe ressaltar ainda que, antes da transformação, nenhuma das variáveis foi considerada normal pelo teste de Shapiro-Wilk realizado. O método de Kolmogorov-Smirnov, por sua vez, reconheceu apenas a variável de Carregamento como regida pela distribuição gaussiana. Entretanto, após a transformação, ambos os testes mencionados confirmaram a normalidade das três variáveis, sendo que, para o Carregamento, houve uma melhoria na estatística do teste de Kolmogorov-Smirnov.

Um outro fator a ser observado é a discrepância do desvio padrão de Embarques quando comparado à sua média. O elevado valor encontrado para essa medida de dispersão, antes da transformação, destaca a ocorrência de eventuais *outliers* no banco de dados e a grande amplitude de variação deste (de 0 a 659 passageiros). A quantidade máxima de Embarques é, possivelmente, um *outlier*, uma vez que se trata do primeiro ponto de parada da linha e deve compreender uma estação de transporte público. É possível afirmar que pontos de parada situados em regiões de alta densidade populacional ou próximos a polos geradores de viagens terão um número de Embarques muito maior que os vizinhos, o que contribui para a elevação do desvio padrão da variável.

Em seguida, foi calculado o índice de Moran para as três variáveis transformadas. A Tabela 4 sintetiza os resultados para esse indicador e sua respectiva significância.

Tabela 4 - Autocorrelação espacial das variáveis de interesse baseada na distância

Variável	Moran's I	Valor Esperado	Padrão	z-score	valor p
Embarques	0,636716	-0,0179	Agrupado	6,4151	0,001
Desembarques	0,304677	-0,0179	Agrupado	3,0643	0,001
Carregamento	0,963610	-0,0182	Agrupado	9,0572	0,001

É importante observar que, além de se ter obtido apenas números maiores que zero para o índice de Moran, o valor *p* resultante desse indicador revela que há uma significativa autocorrelação espacial na estrutura das três variáveis de interesse, já que 0,001 representa o maior nível de significância para 999 permutações (Anselin, 2004). A crescente intensidade no valor

do índice a partir de Desembarques, passando por Embarques e encerrando em Carregamento reflete, também, a dependência espacial presente nessas funções e pode já dar um indício do desempenho do estimador geoestatístico utilizado. Por sua vez, a conclusão encontrada para o padrão revela que, de uma forma geral, as observações próximas entre si tendem a apresentar um comportamento semelhante, condição essencial para o bom desempenho das técnicas de interpolação espacial empreendidas pela Geoestatística.

A Tabela 5 contém as métricas de erros calculadas para os três tipos de semivariogramas considerados. Cabe lembrar que tais resultados se referem aos semivariogramas cujos parâmetros geraram o menor erro quadrado médio, conforme definido na etapa de validação cruzada.

Tabela 5 - Desempenho de três modelos de semivariograma à estimativa das variáveis

	Modelo teórico	SE	ME	MAE	RMSE	R
Embarques	Exponencial	434.845,75	-14,63	34,84	87,34	0,520**
	Gaussiano	435.438,51	-14,89	33,89	87,40	0,527**
	Esférico	433.370,40	-14,69	34,17	87,20	0,528**
Desembarques	Exponencial	142.172,69	-9,72	31,72	49,94	0,418**
	Gaussiano	139.822,02	-10,01	31,86	49,53	0,431**
	Esférico	141.145,33	-9,99	31,88	49,76	0,422**
Carregamento	Exponencial	107.814,30	0,43	27,62	43,88	0,991**
	Gaussiano	85.455,22	-3,95	24,19	39,06	0,993**
	Esférico	96.671,40	0,61	26,24	41,55	0,992**

** A correlação é significativa ao nível de confiança de 99% (1 extremidade).

É importante ressaltar que, apesar de a modelagem dos semivariogramas ótimos ter sido realizada a partir da validação cruzada com os valores das variáveis transformadas, os erros catalogados na Tabela 5 foram obtidos por meio dos valores reais e estimados já resultantes da transformação inversa. Conforme sugerido pelos resultados do índice de Moran, o coeficiente de correlação linear de *Pearson* mostra que, realmente, o desempenho da Krigagem Ordinária melhora de acordo com a sequência: Desembarques, Embarques e Carregamento. Merecem atenção os resultados dessa métrica de aderência e sua significância: embora, nesse quesito, a variável de Carregamento tenha demonstrado um desempenho substancialmente superior ao das demais, todos os valores de R foram bastante significativos, fortalecendo a conclusão de que, certamente, há uma notável semelhança entre os valores reais e seus respectivos estimados pela Krigagem Ordinária.

O ME revela que a maioria das estimativas demonstrou ser inferior ao seu respectivo valor observado, sendo que apenas os semivariogramas exponencial e esférico, para a variável de Carregamento, geraram resultados médios superestimados. Além disso, apesar de haver certa equivalência entre os erros resultantes dos diferentes semivariogramas para cada variável, nota-se que os modelos que tiveram predominância de menores desvios e maiores coeficientes de correlação linear foram o esférico, para o número de Embarques, e o gaussiano, tanto para o

volume de Desembarques quanto para o Carregamento. Tais semivariogramas de melhor desempenho estão mostrados na Figura 3, juntamente com os parâmetros exatos dos três modelos calibrados.

No que tange aos parâmetros mostrados na Figura 3, destacam-se os valores de efeito pepita encontrado para as três variáveis de interesse. Em termos relativos, Embarques demonstrou uma descontinuidade na origem de, aproximadamente, 12,93% do patamar (semivariância máxima, $C_0 + C$), o que certamente contribuiu para um melhor desempenho dessa variável em comparação com Desembarques, cujo efeito pepita foi de 67,86% do patamar. Embora, à primeira vista, essa aleatoriedade seja um fator preocupante à aplicação da Geoestatística, já que excede substancialmente o valor crítico apontado por Landim (2006), ressalta-se que o efeito pepita mínimo dessa variável foi de 39,48%, correspondente ao modelo exponencial e bem mais próximo do limite de 30% do que aquele relativo ao modelo gaussiano. Porém, tendo em vista que o erro quadrado médio para o modelo gaussiano foi inferior ao do exponencial, mantiveram-se os resultados referentes ao semivariograma gaussiano.

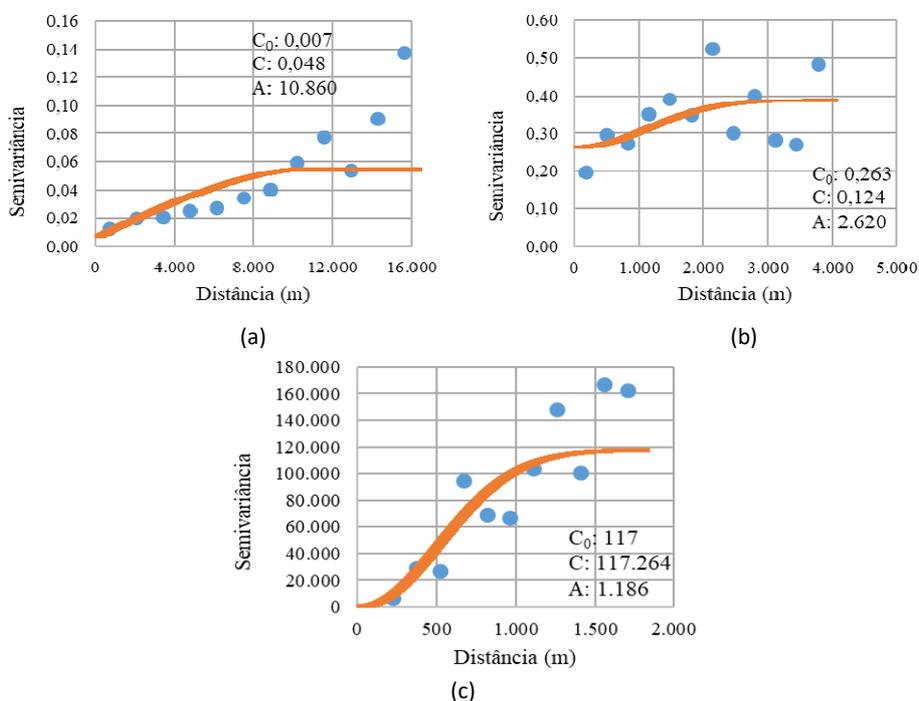


Figura 3. (a) Semivariograma esférico de Embarques; b) Semivariograma gaussiano de Desembarques; e c) Semivariograma gaussiano de Carregamento

A alta descontinuidade na origem concernente ao número de Desembarques por ponto de parada ocorreu, provavelmente, devido a alguma tendência espacial potencialmente presente nessa variável, gerada, por exemplo, por alterações bruscas no ambiente construído do entorno de pontos de parada adjacentes ao longo da linha de ônibus. Esse comportamento pode ser removido do banco de dados, o que envolve uma análise visual subjetiva, ou modelado a partir da inclusão de covariáveis à estrutura da krigagem. Porém, no intuito de avaliar o desempenho da KO isoladamente, tais recursos foram ignorados, permitindo, dessa forma, que abordagens futuras abrangendo esse fator possam gerar resultados passíveis de ser comparados aos da técnica mais simples de interpolação espacial, no âmbito da Geoestatística. Tal verificação também se aplica à variável de Embarques.

É importante destacar que Desembarques apresentou um desempenho inferior ao de Embarques somente com relação ao coeficiente de correlação linear de *Pearson*. Os valores de erro quadrado, erro médio, erro médio absoluto e raiz do erro quadrado médio para Desembarques foram menores que para Embarques, o que mostra uma aderência considerável do volume estimado de passageiros que descem ao real coletado pela pesquisa sobre/desce.

Por fim, destaca-se o excelente desempenho alcançado para a variável Carregamento. Apresentando um efeito pepita desprezível de apenas 0,10% do patamar, suas métricas de erros foram, também, consideravelmente menores que para as outras variáveis, e o coeficiente de correlação linear de *Pearson* resultou quase idêntico à unidade, o que indica uma elevada proximidade entre os valores observados e estimados pela KO.

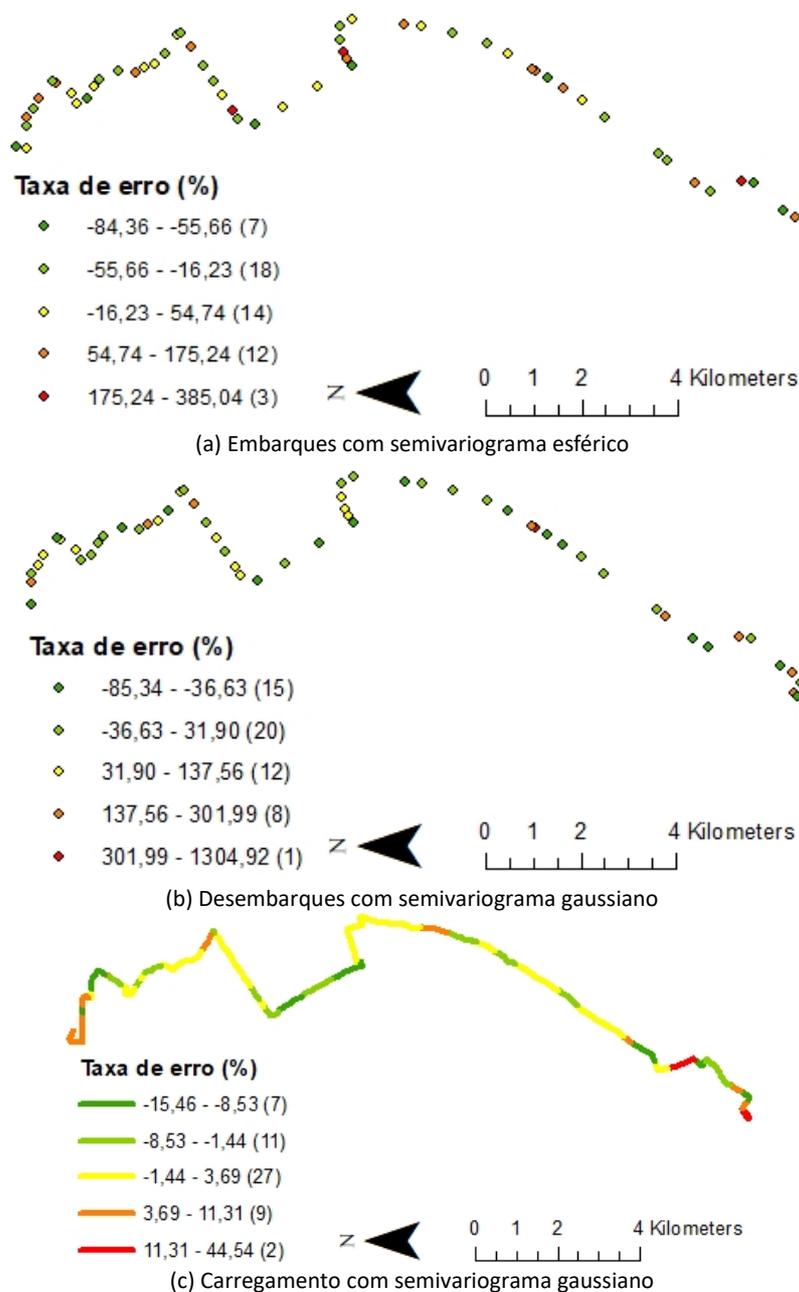


Figura 4. Porcentagem de desvio nas estimativas de krigagem para Embarques, Desembarques e Carregamento

A fim de prover subsídios a uma análise pontual dos erros, os mapas disponibilizados na Figura 4 expõem as taxas de desvio que as estimativas de Embarques, Desembarques e Carregamento, geradas pelos semivariogramas da Figura 3, apresentaram com relação ao valor observado da respectiva variável em cada ponto de parada e trecho de linha. Haja vista que as porcentagens de erro dependem do número real de Embarques, Desembarques e Carregamento, conforme Equação 12, não foi possível calculá-las para os pontos e trechos cujo número de passageiros era nulo. Essa situação foi constatada em três pontos, para Embarques, e apenas um, para Desembarques, os quais não aparecem nos mapas. Carregamento não continha nenhum valor observado nulo.

Um fator que comprova o bom desempenho da KO é a percepção de que a maioria das taxas de desvios se concentra nas categorias intermediárias, para os três casos apresentados na Figura 4: a segunda e terceira faixas de erros (cores verde claro e amarelo, respectivamente) são as que contêm mais pontos da linha. Destaca-se que se tratam das classes que representam os menores desvios, tanto no que se refere à superestimação quanto subestimação. Apenas Desembarques teve uma significativa representatividade do primeiro grupo de erros, o que corrobora o menor coeficiente de correlação entre valores estimados e observados dessa variável em comparação com as demais e evidencia, novamente, sua substancial aleatoriedade espacial.

Conhecendo o perfil das três variáveis de interesse, é possível afirmar que a ocorrência de erros extremos, no caso de Embarques e Desembarques, é devida ao fato de tais funções não serem tão suaves, espacialmente, quanto Carregamento. A sobreposição dos mapas da Figura 4 com os da Figura 2 revela que as maiores taxas de desvio, que contêm três representantes em Embarques e um em Desembarques, ocorrem em pontos de parada com movimentação atípica de passageiros se comparados aos pontos vizinhos. Uma vez que os pontos vizinhos apresentavam um volume de Embarques e de Desembarques substancialmente maior, a krigagem previu, para esses *outliers* locais, um valor estimado muito maior que o real. Embarques e Desembarques apresentaram variações bruscas na vizinhança próxima que a Krigagem Ordinária não conseguiu modelar satisfatoriamente, já que essa ferramenta exige que as observações das variáveis de estudo sejam semelhantes entre si a pequenas distâncias.

Tal problema pode ser resolvido a partir da adição de variáveis independentes à estrutura do estimador, tais como densidade populacional, de empregos, uso do solo, acessibilidade, entre outras, que expliquem o motivo para a presença das variações abruptas supracitadas. *Outliers* também podem ser controlados por meio de estimadores robustos do semivariograma empírico, como o proposto por Cressie e Hawkins (1958). Além disso, se os dados não são normalizados, o ajuste do semivariograma teórico pode ser feito pelo método de mínimos quadrados generalizados, que contabiliza a variância não constante da variável de interesse.

A exemplo das variáveis de Embarques e Desembarques, Carregamento também demonstra uma predominância de erros nas categorias intermediárias, ou seja, a maioria dos desvios se refere a pequenas variações para mais ou para menos do valor real. Entretanto, quando comparado às duas primeiras variáveis, é notável o quanto Carregamento exhibe um excelente desempenho: os erros se mantêm entre -8% e 4%, fato que, associado à quantidade irrisória de informações necessárias à modelagem, consolida definitivamente a adequabilidade do ferramental geoestatístico à estimativa da demanda por transporte público ao longo de linhas de ônibus.

As taxas de desvio de Embarques, Desembarques e Carregamento apresentaram uma assimetria positiva também verificada em resultados de trabalhos anteriores (Chi e Zheng, 2013; Wang e Kockelman, 2009). A mediana dos erros em porcentagem foi de 6,59%, -3,36% e -0,08%,

respectivamente, para Embarques, Desembarques e Carregamento nos casos da Figura 4, o que confirma, mais uma vez, a tendência de subestimação do estimador de krigagem. Em se tratando dos desvios absolutos, as três variáveis demonstraram uma mediana de aproximadamente 39%, 54% e 3%, e média de 69%, 95% e 5%. Apesar de os resultados obtidos para Embarques e Desembarques parecerem desanimadores, em comparação com estudos que utilizaram a KO com validação cruzada (Chi e Zheng, 2013; Shamo *et al.*, 2015), tais valores podem ser reconhecidos como satisfatórios. A média de erros encontrada por Chi e Zheng (2013) ficou em torno de 59%, porém, o menor valor de RMSE calculado por esses autores foi de 124. Cabe lembrar que o maior RMSE obtido no presente trabalho é da ordem de 87, conforme mostrado na Tabela 5. Já Shamo *et al.* (2015) obtiveram RMSEs de 56, 83 e 95, enquanto que o menor valor desse erro, para Embarques, Desembarques e Carregamento, respectivamente, foi de 87, 49 e 39.

Por fim, conforme definido na subseção 2.5., as técnicas de interpolação espacial da Geoestatística originam superfícies contínuas de valores estimados. Dessa forma, tanto os pontos amostrados em campo quanto aqueles que não foram contemplados pelo levantamento de dados recebem informações relativas às variáveis de interesse. A Figura 5 exemplifica esse procedimento ao mostrar, à direita, o número de Embarques atribuído a pontos não amostrados (em vermelho), coletado a partir da superfície contínua à esquerda, que corresponde à krigagem de Embarques com semivariograma esférico.

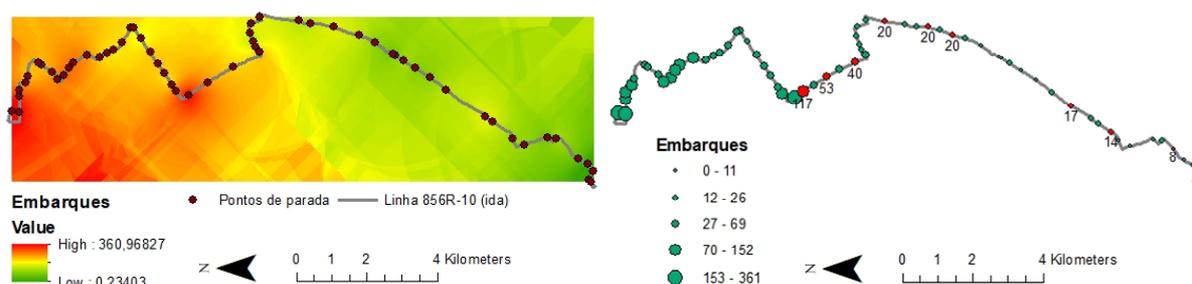


Figura 5: Estimativa de Embarques em pontos não amostrados

4. CONCLUSÕES E CONSIDERAÇÕES FINAIS

O presente trabalho se valeu da dependência espacial entre as observações de viagens no intuito de gerar estimativas do volume de Embarques e Desembarques, por ponto de parada, e Carregamento nos trechos, ao longo de uma linha de transporte público por ônibus, que não seriam amostrados por ocasião da pesquisa sobe/desce, a fim de desobrigar a realização desse levantamento em toda a rede de TP. Tal previsão foi realizada por meio da Krigagem Ordinária, técnica de interpolação geoestatística que, para sua aplicação, exige apenas o valor da variável em pontos do espaço e suas respectivas coordenadas geográficas.

É possível afirmar que o estudo de caso aplicado à linha 856R-10 forneceu um vislumbre da eficácia da Geoestatística em prever o volume de passageiros ao longo de uma linha de transporte público. Embora ainda não se possa generalizar os resultados, as métricas de aderência revelaram que, mesmo requerendo um número mínimo de informações para a realização da modelagem geoestatística, a Krigagem Ordinária conseguiu alcançar ótimas estimativas.

Essa afirmação ganha maiores proporções em se tratando, especialmente, de Carregamento: devido ao fato de ser a mais homogênea dentre as três variáveis de interesse, observou-se que

a maior parte dos erros dos valores previstos por trecho se manteve na pequena faixa de -8% a 4% do volume real de passageiros. Tal resultado representa um importante avanço no planejamento do transporte público por ônibus, tendo em vista que o Carregamento no segmento crítico, aquele em que o número de usuários dentro do veículo alcança sua ocupação máxima ao longo da viagem, é justamente um dos parâmetros utilizados para um dimensionamento otimizado da frota e, conseqüentemente, elaboração do cronograma de partidas.

Um dos motivos para a obtenção dos bons resultados observados no presente trabalho pode ser atribuído ao nível de agregação das variáveis de interesse e homogeneidade no tamanho e forma dos suportes utilizados (pontos de parada e trechos de linha). Conforme mencionado na primeira seção deste texto, essa condição ainda não havia sido verificada em trabalhos anteriores (Gomes *et al.*, 2016; Lindner e Pitombo, 2017; Pitombo *et al.*, 2015) e pode ter representado fator contribuinte para uma melhoria na modelagem geoestatística do presente trabalho em comparação com os supracitados.

Cabe destacar, também, que os resultados alcançados nesse trabalho dão suporte tanto à modelagem geoestatística de redes de TP que já dispõem das informações de Embarques, Desembarques e Carregamento (coletadas a partir de uma pesquisa sobre/desce ou outra tecnologia, como os contadores automáticos de passageiros) quanto daquelas nas quais uma pesquisa sobre/desce ainda será realizada. No primeiro caso, a krigagem pode ser utilizada para o conhecimento prévio do número de usuários que poderão embarcar ou desembarcar em um novo ponto de parada a ser implantado ao longo das linhas de ônibus. Além disso, a superfície de interpolação gerada pelo estimador geoestatístico permite que alterações na demanda, provenientes de eventuais mudanças no traçado das rotas, sejam avaliadas e norteiem decisões quanto à programação das viagens e ajustes na oferta. No segundo cenário, por sua vez, a Geoestatística ainda permite que a pesquisa de Embarque e Desembarque não seja realizada em todos os trechos das linhas de ônibus nem em toda a rede de transporte público, facilitando sua execução haja vista a redução nas despesas inerentes a esse levantamento.

Dessa forma, o presente trabalho contribui para a disseminação do planejamento de redes de transporte público pautado em levantamentos de Embarque e Desembarque menos onerosos que os tradicionais. Ademais, a investigação aqui proposta introduz, com sucesso, a viabilidade de aplicação da Geoestatística às variáveis de demanda por transporte público ao longo de linhas de ônibus, objetos de interesse do planejamento de redes de TP. Porém, por brevidade, foi possível mostrar apenas a técnica de interpolação geoestatística mais simplificada. Dessa forma, pretende-se confrontar, em trabalhos futuros, o desempenho da Krigagem Ordinária ao da Krigagem Universal, cuja introdução de covariáveis à estrutura do estimador pode levar a resultados ainda melhores. A Regressão Geograficamente Ponderada, que não supõe continuidade espacial do fenômeno estudado e também possui variáveis explicativas em sua formulação, surge como uma alternativa promissora a ser comparada à Krigagem Ordinária. Além disso, buscar-se-á, também, identificar o conjunto de pontos de parada e trechos de linhas, a serem contemplados pela pesquisa sobre/desce, que gerariam superfícies contínuas de valores estimados mais próximas da condição em que as métricas de aderência são otimizadas.

AGRADECIMENTOS

Às agências de fomento CAPES, CNPq (Processo 427525/2018-7; 303645/2015-6) e FAPESP (Processo 13/25035-1). Os autores também agradecem à SPTrans, pela disponibilização da pesquisa de Embarque e Desembarque utilizada nesse trabalho.

REFERÊNCIAS

- Anselin, L. (2004) Exploring spatial data with GeoDaTM: a workbook. *Urbana*, vol. 51, n. 61801. Disponível em: <http://www.csiss.org/clearinghouse/GeoDa/geodaworkbook.pdf>. Acesso em: ago. 2019.
- Anselin, L., I. Syabri e Y. Kho (2005) GeoDa : An Introduction to Spatial Data Analysis. *Geographical Analysis*, vol. 38, n. 1, p. 5–22. <https://doi.org/10.1111/j.0016-7363.2005.00671.x>
- Bagchi, M. e P. R. White (2005) The potential of public transport smart card data. *Transport Policy*, v. 12, n. 5, p. 464–474, 2005. <https://doi.org/10.1016/j.tranpol.2005.06.008>
- Barbosa, H. M., C. M. P. Braga, J. A. P. Coelho, T. Arruda e F. G. Jota (2006) Sistema Automático de Monitoração da Movimentação de Passageiros do Transporte Coletivo por Ônibus - Um Projeto Integrado de Ensino e Desenvolvimento Tecnológico. *XX Congresso Nacional de Pesquisa em Transporte da Anpet*. Disponível em: https://www.researchgate.net/publication/281441109_Sistema_Automatico_de_Monitoracao_da_Movimentacao_de_Passageiros_do_Transporte_Coletivo_por_Onibus. Acesso em: ago. 2019.
- Bartlett, M. S. (1947) The Use of Transformations. *Biometrics*, vol. 3, n. 1, p. 39–52. <https://doi.org/10.2307/3001536>
- Bayraktar, H. e F. S. Turalioglu (2005) A Kriging-based approach for locating a sampling site in the assessment of air quality. *Stochastic Environmental Research and Risk Assessment*, vol. 19, n. 4, p. 301–305. <https://doi.org/10.1007/s00477-005-0234-8>
- Box, G. E. P. e D. R. Cox (1964) An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 26, n. 2, p. 211–252. Disponível em: <https://www.jstor.org/stable/2984418>. Acesso em: ago. 2019. <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>
- Ceder, A. (2007) *Public transit planning and operation: modeling, practice and behavior*. CRC press. <https://doi.org/10.1201/b12853>
- Chi, G., e Zheng, Y. (2013) Estimating Transport Footprint along Highways at Local Levels: A Combination of Network Analysis and Kriging Methods. *International Journal of Sustainable Transportation*, vol. 7, n. 3, p. 261–273. <https://doi.org/10.1080/15568318.2013.710150>
- Cressie, N. (1993) *Statistics for spatial data*. John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119115151>
- Cressie, N. (1985) Fitting variogram models by weighted least squares. *Journal of the International Association for Mathematical Geology*, vol. 17, n. 5, p. 563–586. DOI:10.1007/BF01032109
- Cressie, N. e D. M. Hawkins (1980). Robust estimation of the variogram. *I. Journal of the International Association for Mathematical Geology*, vol. 12, n. 2, p. 115-125. <https://doi.org/10.1007/BF01035243>
- Eom, J. I. N. K. I., S. I. K. P. Man, T.-Y. Heo e L. F. Huntsinger (2006) Improving the prediction of annual average daily traffic for nonfreeway facilities by applying a spatial statistical method. *Transportation Research Record*, n. 1968, p. 22–29. <https://doi.org/10.3141/1968-03>
- ESRI (2010) Geostatistical Analyst Tutorial. Environmental Systems Research Institute. Disponível em: <<http://help.arcgis.com/en/arcgisdesktop/10.0/pdf/geostatistical-analyst-tutorial.pdf>> Acesso em: mai. 2018.
- Ferraz, A. C. P. e I. G. E. Torres (2004) *Transporte público urbano*. (2º ed). RiMa Editora.
- Gomes, M. M., C. S. Pitombo, A. Pirdavani e T. Brijs (2018) Geostatistical approach to estimate car occupant fatalities in traffic accidents. *Revista Brasileira de Cartografia*, vol. 70, n. 4, p. 1231–1256. <https://doi.org/10.14393/rbcv70n4-46140>
- Gomes, V. A., C. S. Pitombo, S. S. Rocha e A. R. Salgueiro (2016) Kriging Geostatistical Methods for Travel Mode Choice: A Spatial Data Analysis to Travel Demand Forecasting. *Open Journal of Statistics*, vol. 6, n. 3, p. 514–527. <https://doi.org/10.4236/ojs.2016.63044>
- Goovaerts, P. (2008) Kriging and Semivariogram Deconvolution in the Presence of Irregular Geographical Units. *Mathematical geology*, vol. 40, n. 1, p. 101–128. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2518693/pdf/nihms-36655.pdf>. Acesso em: ago. 2019. <https://doi.org/10.1007/s11004-007-9129-1>
- Goovaerts, P. (2012) Geostatistical Analysis of Health Data with Different Levels of Spatial Aggregation. *Spatial and Spatio-Temporal Epidemiology*, vol. 3, n. 1, p. 83–92. <https://doi.org/10.1016/j.sste.2012.02.008>
- Gundogdu, I. B. (2014) Risk governance for traffic accidents by Geostatistical Analyst methods. *International Journal of Research in Engineering and Science*, v. 2, n. 9, p. 35-40. Disponível em: <http://www.ijres.org/papers/Volume%202/v2-i9/D293540.pdf>. Acesso em: ago. 2019.
- Hollander, Y. e R. Liu (2008) The principles of calibrating traffic microsimulation models. *Transportation*, vol. 35, n. 3, p. 347–362. <https://doi.org/10.1007/s11116-007-9156-2>
- IBGE (2018) São Paulo. IBGE Cidades, Instituto Brasileiro de Geografia e Estatística. Disponível em: <<https://cidades.ibge.gov.br/brasil/sp/sao-paulo/panorama>>. Acesso em: mai. 2018.
- IBM (2016) IBM SPSS Statistics 24 Core System User's Guide. International Business Machines. Disponível em: <ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/24.0/en/client/Manuals/IBM_SPSS_Statistics_Core_System_User_Guide.pdf> Acesso em: mai. 2018.
- 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. <https://doi.org/10.1061/JTEPBS.0000064>

- Krige, D. G. (1951) A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 52, n. 6, p. 119–139. Disponível em: <https://journals.co.za/docserver/fulltext/saimm/52/6/4823.pdf?expires=1566996803&id=id&accname=guest&checksum=FF6F1BA4B54466840936B8253F9710CA>. Acesso em: ago. 2019.
- Landim, P. M. B. (2006) Sobre geoestatística e mapas. *Terrae Didactica*, vol. 2, n. 1, p. 19–33. <https://doi.org/10.20396/td.v2i1.8637463>
- Lindner, A. e C. S. Pitombo (2017) A Conjoint Approach of Spatial Statistics and a Traditional Method for Travel Mode Choice Issues. *Journal of Geovisualization and Spatial Analysis*, vol. 2, n. 1. <https://doi.org/10.1007/s41651-017-0008-0>
- Lindner, A., C. S. Pitombo, S. S. Rocha e J. A. Quintanilha (2016) Estimation of transit trip production using Factorial Kriging with External Drift: an aggregated data case study. *Geo-spatial Information Science*, vol. 19, n. 4, p. 245–254. <https://doi.org/10.1080/10095020.2016.1260811>
- Majumdar, A., R. B. Noland e W. Y. Ochieng (2004) A spatial and temporal analysis of safety-belt usage and safety-belt laws. *Accident Analysis & Prevention*, vol. 36, n. 4, p. 551–560. [https://doi.org/10.1016/S0001-4575\(03\)00061-7](https://doi.org/10.1016/S0001-4575(03)00061-7)
- Manepalli, U. R. R. e G. H. Bham (2011) Crash prediction: evaluation of empirical Bayes and kriging Methods. *3rd International Conference on Road Safety and Simulation*. Disponível em: <http://onlinepubs.trb.org/onlinepubs/conferences/2011/RSS/1/Manepalli,U.pdf>. Acesso em: ago. 2019.
- Marques, S. F. (2019) Estimativa do volume de passageiros ao longo de uma linha de transporte público por ônibus a partir da Geoestatística. 98 p. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2019. <https://doi.org/10.11606/D.18.2019.tde-26042019-110232>.
- Matheron, G. (1971) The Theory of Regionalized Variables and Its Applications. Les Cahiers du Centre de Morphologie Mathématique in Fontainebleu, Paris.
- Millard, S. P. (2013) *EnvStats*. (Springer, Ed) *Springer eBooks*. vol. 2. <https://doi.org/10.1007/978-1-4614-8456-1>
- Moran, P. A. P. (1948) The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 10, n. 2, p. 243–251. Disponível em: <https://www.jstor.org/stable/2983777>. Acesso em: ago. 2019. <https://doi.org/10.1111/j.2517-6161.1948.tb00012.x>
- Olea, R. A. (2006) A six-step practical approach to semivariogram modeling. *Stochastic Environmental Research and Risk Assessment*, vol. 20, n. 5, p. 307–318. <https://doi.org/10.1007/s00477-005-0026-1>
- Oliver, M. A. e R. Webster (1986) Combining Nested and Linear Sampling for Determining the Scale and Form of Spatial Variation of Regionalized Variables. *Geographical Analysis*, vol. 18, n. 3, p. 227–242. <https://doi.org/10.1111/j.1538-4632.1986.tb00095.x>
- Oliver, M. A. e R. Webster (2015) *Basic steps in geostatistics: the variogram and kriging*. Springer. <https://doi.org/10.1007/978-3-319-15865-5>
- Pitombo, C. S., A. R. Salgueiro, A. S. G. da Costa e C. A. Isler (2015) A two-step method for mode choice estimation with socioeconomic and spatial information. *Spatial Statistics*, vol. 11, p. 45–64. <https://doi.org/10.1016/j.spasta.2014.12.002>
- R Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rocha, S. S., A. Lindner e C. S. Pitombo (2017) Proposal of a Geostatistical Procedure for Transportation Planning Field. *Boletim de Ciências Geodésicas*, v. 23, n. 4, p. 636–653. <https://doi.org/10.1590/s1982-21702017000400042>.
- São Paulo (2015) Plano de Mobilidade de São Paulo. Prefeitura do Município de São Paulo, Secretaria Municipal de Transportes. Disponível em: <http://www.prefeitura.sp.gov.br/cidade/secretarias/upload/chamadas/planmobsp_v072_1455546429.pdf>. Acesso em: mai. 2018.
- Selby, B. e K. M. Kockelman (2013) Spatial prediction of traffic levels in unmeasured locations: Applications of universal kriging and geographically weighted regression. *Journal of Transport Geography*, vol. 29, p. 24–32. <https://doi.org/10.1016/j.jtrangeo.2012.12.009>
- Shamo, B., Asa, E., e Membah, J. (2015) Linear Spatial Interpolation and Analysis of Annual Average Daily Traffic Data. *Journal of Computing in Civil Engineering*, vol. 29, n. 1, 4014022. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000281](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000281)
- Shapiro, S. S., e Wilk, M. B. (1965) An analysis of variance test for normality (complete samples). *Biometrika*, vol. 52, n. 3–4, p. 591–611. <https://doi.org/10.1093/biomet/52.3-4.591>
- SPTrans (2018) Dados SPTrans 2017. São Paulo Transporte S.A.
- Subhadip, B., C. Souvik, C. Satish e G. Indrajit (2017) Kriging-Based Approach for Estimation of Vehicular Speed and Passenger Car Units on an Urban Arterial. *Journal of Transportation Engineering, Part A: Systems*, vol. 143, n. 3. <https://doi.org/10.1061/JTEPBS.0000031>
- Vogel, R. M. (1986) The Probability Plot Correlation Coefficient Test for the Normal, Lognormal, and Gumbel Distributional Hypotheses. *Water Resources Research*, vol. 22, n. 4, p. 587–590. <https://doi.org/10.1029/WR022i004p00587>
- Vuchic, V. R. (2005) *Urban Transit: Operations, Planning, and Economics*. John Wiley & Sons.
- Wang, X. e K. Kockelman (2009) Forecasting Network Data. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2105, p. 100–108. <https://doi.org/10.3141/2105-13>
- Yamamoto, J. K. e P. M. B. Landim (2015) *Geoestatística: conceitos e aplicações*. Oficina de textos.
- Zhang, D. e X. C. Wang (2014) Transit ridership estimation with network Kriging: A case study of Second Avenue Subway, NYC. *Journal of Transport Geography*, vol. 41, p. 107–115. <https://doi.org/10.1016/j.jtrangeo.2014.08.021>.

Zimmerman, D. A., G. de Marsily, C. A. Gotway, M. G. Marietta, C. L. Axness, R. L. Beauheim, R. L. Bras, J. Carrera, G. Dagan e P. B. Davies (1998) A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow. *Water Resources Research*, vol. 34, n. 6, p. 1373-1413.
<https://doi.org/10.1029/98WR00003>