

ESTIMATIVA DO VOLUME DE PASSAGEIROS AO LONGO DE LINHAS DE TRANSPORTE PÚBLICO POR ÔNIBUS A PARTIR DA GEOESTATÍSTICA

Samuel de França Marques

Cira Souza Pitombo

Universidade de São Paulo

Escola de Engenharia de São Carlos

RESUMO

A pesquisa sobre/desce desempenha um importante papel no planejamento de redes de transporte público. Contudo, esse levantamento demanda elevados recursos financeiros e de tempo, além de uma equipe de profissionais bem treinados para a visita das linhas de ônibus. Dessa forma, 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 linhas de transporte público, que não seriam amostrados por ocasião da pesquisa. 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. Os resultados obtidos, a partir da validação cruzada, apontaram um ajuste considerável dos valores estimados aos observados, confirmando a adequabilidade do ferramental geoestatístico à estimativa de variáveis de demanda por transportes.

ABSTRACT

The boarding and alighting counts survey plays an important role in planning public transport networks. However, this survey demands high financial and time resources, as well as a team of well-trained professionals to visit bus lines. In this way, the present study relies on the spatial dependence between the observations of trips in order to generate estimates of the volume of boarding and alighting per bus stop and loading in the stretches, along public transport lines, that would not be sampled for the survey occasion. Such prediction was performed by ordinary kriging, a geostatistical interpolation technique that, for its application, only requires the value of the variable in points of the space and their respective geographical coordinates. The results obtained, from the cross-validation, indicated a considerable adjustment of the estimated values to the observed ones, confirming the suitability of the geostatistical tool to the estimation of travel demand variables.

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 sobre/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, consequentemente, elaboração do cronograma de viagens, descrita por Ferraz e Torres (2004), exige a utilização do índice de renovação, fator computado a partir da razão entre o total de passageiros transportados e o volume na seção de maior carregamento, dados que só podem ser obtidos 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 visita 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, ferramenta 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, empregadas pela Geoestatística, permitiriam que, a partir da demanda referente apenas a determinados pontos de parada ou trechos de linhas de ônibus visitados por ocasião da pesquisa de embarque e desembarque, fosse possível encontrar os valores dessa variável nos pontos ou 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 fez com que seus conceitos fossem expandidos a diversas áreas de estudo (Bayraktar e Turalioglu, 2005; Goovaerts, 2012; Oliver e Webster, 2010; Zimmerman *et al.*, 1998). Na engenharia de transportes, o emprego dos conceitos geoestatísticos já foi verificado na modelagem de acidentes/segurança viária (Majumdar *et al.*, 2004) 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 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 (Benedict *et al.*, 2015; Eom *et al.*, 2006; Klatko *et al.*, 2017; Selby e Kockelman, 2013; 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 agregada 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, 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 apenas por meio da pesquisa sobre/desce ou a partir de contadores automáticos de passageiros (Barbosa *et al.*, 2006), que, devido ao seu elevado custo, ainda não foram popularizados.

Dessa forma, o objetivo 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, conforme descrição apresentada no item 2.

2. MATERIAIS E MÉTODO

A seguir, introduz-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 fazia necessária a aquisição de um banco de dados proveniente de uma pesquisa sobre/desce, na qual fossem identificados os pontos de parada e suas respectivas coordenadas geográficas. Foram disponibilizados, 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 desembarque 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.

A partir de uma inspeção visual do comportamento espacial das oito linhas com os dados de 2017, a linha 856R-10 foi escolhida para compor o estudo de caso por apresentar uma estrutura espacial aparente em seu perfil de movimentação de passageiros, que poderia gerar

resultados satisfatórios tão somente por meio da aplicação da técnica de interpolação espacial mais simples, convergindo com o objetivo do presente trabalho em prover uma análise piloto do desempenho da Geoestatística. 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). Em suma, consolidou-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. 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.

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; 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.

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, que se torna ainda mais evidente em se tratando da demanda por transporte público, 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).

Após essa operação, as variáveis resultantes da transformação de Box-Cox passaram por uma etapa de análise espacial exploratória, 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 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 2.

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

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 2 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 (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. Tendo em vista que a linha 856R-10 se estende por uma orientação quase bem definida, é possível afirmar que, nesse caso, o semivariograma omnidirecional traz respostas semelhantes às do semivariograma que considera a direção aproximada da linha, cuja identificação demandaria esforços de programação. Ademais, o semivariograma omnidirecional consegue abranger uma maior quantidade de pares, o que melhora a precisão dos resultados.

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 podem 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), 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 o exponencial, gaussiano e esférico.

O ajuste do semivariograma empírico a esses três modelos teóricos será efetuado por meio da minimização dos mínimos quadrados ponderados (WLS), método proposto por Cressie (1985). Em se tratando dos mínimos quadrados ponderados, as semivariâncias médias calculadas por meio da Equação 2 recebem pesos de acordo com o número de pares encontrados a cada distância h . 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.

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. 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 1.

Tabela 1: Métricas de aderência

Nome	Equação	Comentários
Erro Quadrado (SE)	$\sum_{i=1}^n [y(s_i) - y_{-i}^*(s_i)]^2 \quad (3)$	-
Erro Médio (ME)	$\frac{1}{n} \sum_{i=1}^n [y_{-i}^*(s_i) - y(s_i)] \quad (4)$	Indica a existência de erros sistemáticos
Erro Médio Absoluto (MAE)	$\frac{1}{n} \sum_{i=1}^n y(s_i) - y_{-i}^*(s_i) \quad (5)$	Não particularmente sensitivo a grandes erros
Raiz Quadrada do Erro Quadrado Médio (RMSE)	$\sqrt{\frac{1}{n} \sum_{i=1}^n [y(s_i) - y_{-i}^*(s_i)]^2} \quad (6)$	Grandes erros são duramente penalizados
Coefficiente de correlação linear de Pearson (R)	$\frac{1}{n} \sum_{i=1}^n \frac{[y(s_i) - \bar{y}][y_{-i}^*(s_i) - \bar{y}^*]}{\sigma_y \sigma_{y^*}} \quad (7)$	-

Fonte: Adaptada de Hollander e Liu (2008)

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 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 1, 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 e 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 aleatória 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 8.

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (8)$$

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 devido a 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) de que o somatório dos pesos deve ser igual a 1 (Yamamoto e Landim, 2015). Dessa forma, além de permitir inferência estatística nas estimativas, recurso ausente em técnicas de interpolação espaciais 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 8 e as duas restrições mencionadas anteriormente, estabelece o sistema de equações da krigagem ordinária em termos da função semivariograma, conforme mostrado na Equação 9 (Yamamoto e Landim, 2015).

$$\begin{cases} \sum_{j=1}^n \lambda_j \gamma(x_i - x_j) + \mu = \gamma(x_0 - x_i) & \text{para } i = 1, n \\ \sum_{j=1}^n \lambda_j = 1 \end{cases} \quad (9)$$

Em que μ corresponde à média local da variável regionalizada. 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 Tabela 2 consolida as medidas de estatística descritiva tanto para as variáveis de interesse em sua forma bruta quanto transformadas pelo método de Box-Cox.

Tabela 2: 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 um acréscimo na estatística do teste de Kolmogorov-Smirnov. Em seguida, foi calculado o índice de Moran para as três variáveis transformadas. A Tabela 3 sintetiza os resultados para esse indicador e sua respectiva significância.

Tabela 3: Autocorrelação espacial das variáveis de interesse

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 4 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 4: 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 4 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 1 juntamente com os parâmetros exatos dos três modelos calibrados.

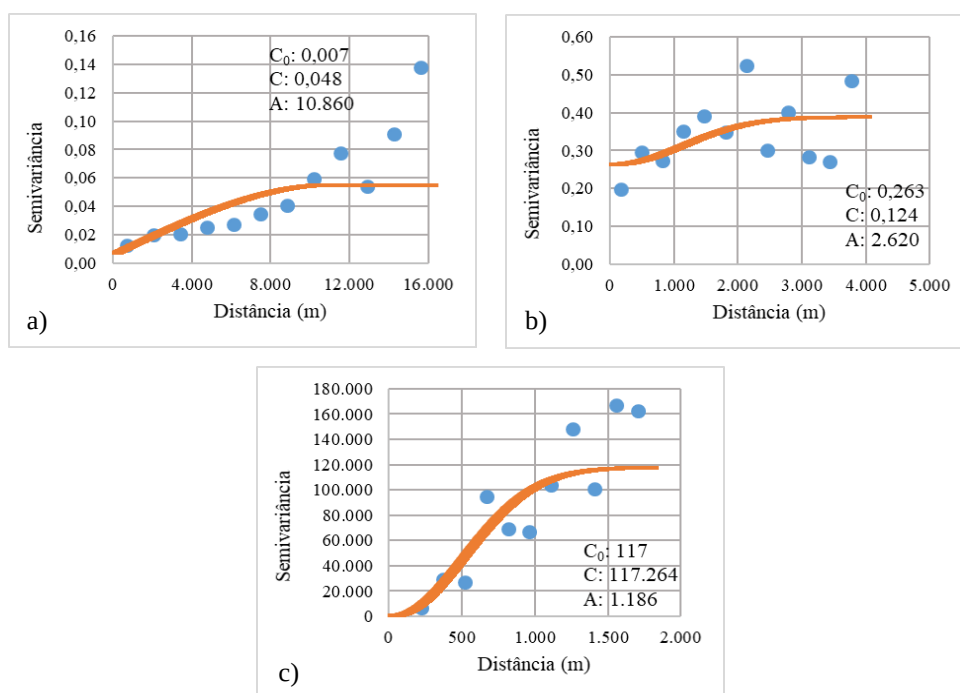


Figura 1: a) Semivariograma esférico de Embarques; b) Semivariograma gaussiano de Desembarques; e c) Semivariograma gaussiano de Carregamento

No que tange aos parâmetros mostrados na Figura 1, destaca-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 à de 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 segundo modelo foi inferior ao do primeiro, mantiveram-se os resultados referentes ao semivariograma gaussiano.

A alta descontinuidade na origem concernente ao número de desembarques por ponto de parada ocorreu, provavelmente, devido à 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 serem comparados aos da técnica mais simples de interpolação espacial, no âmbito da Geoestatística. Tal verificação se aplica também à variável de embarques.

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.

4. CONCLUSÕES E CONSIDERAÇÕES FINAIS

A pesquisa sobre/desce desempenha um importante papel no planejamento de redes de transporte público. Contudo, esse levantamento demanda elevados recursos financeiros e de tempo, além de uma equipe de profissionais bem treinados para a visita das linhas de ônibus. Dessa forma, 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 linhas de transporte público, que não seriam amostrados por ocasião da pesquisa, 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. Os resultados apontaram um ajuste considerável dos valores estimados aos observados, confirmando a adequabilidade do ferramental geoestatístico à estimativa de variáveis de demanda por transportes.

Cabe destacar, também, que os resultados alcançados nesse trabalho dão suporte tanto à modelagem geoestatística de redes de TP que já dispõem de uma pesquisa sobre/desce completa quanto daquelas nas quais esse levantamento ainda será realizado. 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. A

Figura 2 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. 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.

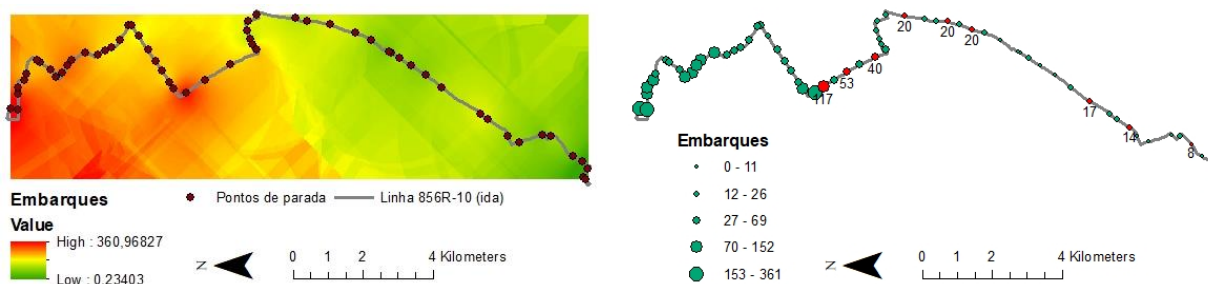


Figura 2: Estimativa de embarques em pontos não amostrados

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. Além disso, buscar-se-á, também, identificar o conjunto de pontos de parada e trechos de linhas, a serem contemplados pela pesquisa sobe/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 e FAPESP (Processo 13/25035-1). Os autores também agradecem à SPTrans, pela disponibilização da pesquisa de embarque e desembarque utilizada nesse trabalho.

REFERÊNCIAS BIBLIOGRÁFICAS

- Anselin, L. (2004) Exploring spatial data with GeoDaTM: a workbook. *Urbana*, vol. 51, n. 61801.
- Anselin, L., I. Syabri e Y. Kho (2005) GeoDa : An Introduction to Spatial Data Analysis. *Geographical Analysis*, vol. 38, n. 1, p. 5–22. doi:10.1111/j.0016-7363.2005.00671.x
- 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*.
- 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.
- Benedict, S., A. Eric e M. Joseph (2015) Linear Spatial Interpolation and Analysis of Annual Average Daily Traffic Data. *Journal of Computing in Civil Engineering*, vol. 29, n. 1.
- 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.
- 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
- 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. doi:10.3141/1968-03

- Ferraz, A. C. P. e I. G. E. Torres (2004) *Transporte público urbano*. (2º ed). RiMa Editora.
- 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. doi: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.
- 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. doi:10.1016/j.sste.2012.02.008
- Hollander, Y. e R. Liu (2008) The principles of calibrating traffic microsimulation models. *Transportation*, vol. 35, n. 3, p. 347–362. doi:10.1007/s11116-007-9156-2
- 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
- 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.
- Landim, P. M. B. (2006) Sobre geoestatística e mapas. *Terrae Didactica*, vol. 2, n. 1, p. 19–33.
- 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.
- 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. doi: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. doi:https://doi.org/10.1016/S0001-4575(03)00061-7
- Matheron, G. (1971) The Theory of Regionalized Variables and Its Applications. Les Cahiers du Centre de Morphologie Mathématique in Fontainebleau, Paris.
- Millard, S. P. (2013) *EnvStats*. (Springer, Ed) Springer eBooks. vol. 2.
- 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.
- Oliver, M. A. e R. Webster (2010) 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.
- 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. doi: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.
- 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. doi:10.1016/j.jtrangeo.2012.12.009
- 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. doi: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.
- 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. doi: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. doi: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.