UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE SÃO CARLOS

SAMUEL DE FRANÇA MARQUES

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

São Carlos
2019

SAMUEL DE FRANÇA MARQUES

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

Programa de Pós-graduação em
Engenharia de Transportes da EESC-USP

**Exemplar definitivo (corrigido). O exemplar original está disponível na CPG da EESC-USP**

São Carlos, **22/04/2019**

**RESOLUÇÃO CoPGr N º 6018, DE 13 DE OUTUBRO DE 2011, artigo 5º**

Dissertação apresentada à Escola de Engenharia de São Carlos da Universidade de São Paulo, como parte dos requisitos para a obtenção do Título de Mestre em Ciências.

Orientador: Prof. Dr. Cira Souza Pitombo

São Carlos

2019

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da EESC/USP com os dados inseridos pelo(a) autor(a).

D187e

DE FRANÇA MARQUES, SAMUEL
ESTIMATIVA DO VOLUME DE PASSAGEIROS AO LONGO DE UMA LINHA DE TRANSPORTE PÚBLICO POR ÔNIBUS A PARTIR DA GEOESTATÍSTICA / SAMUEL DE FRANÇA MARQUES; orientadora CIRA SOUZA PITOMBO. São Carlos, 2019.

Dissertação (Mestrado) - Programa de Pós-Graduação em Engenharia de Transportes e Área de Concentração em Planejamento e Operação de Sistemas de Transporte -- Escola de Engenharia de São Carlos da Universidade de São Paulo, 2019.

1. Estatística espacial. 2. Pesquisa sobe/desce. 3. Krigagem. 4. Demanda por transporte público. 5. Distâncias em rede. 6. Distâncias euclidianas. I. Título.

Eduardo Graziosi Silva - CRB - 8/8907

# FOLHA DE JULGAMENTO

Candidato: Bacharel **SAMUEL DE FRANÇA MARQUES.**

Título da dissertação: "Estimativa do volume de passageiros ao longo de uma linha de transporte público por ônibus a partir da Geoestatística".

Data da defesa: 15/02/2019.

| **Comissão Julgadora:** | **Resultado:** |
|---|---|
| Profa. Associada **Cira Souza Pitombo** **(Orientadora)** (Escola de Engenharia de São Carlos/EESC) | Aprovado |
| Prof. Dr. **Bruno Vieira Bertoncini** (Universidade Federal do Ceará/UFC) | Aprovado |
| Profa. Dra. **Ana Ria Gonçalves Neves Lopes Salgueiro** (Universidade Federal do Ceará/UFC) | Aprovado |

Coordenadora do Programa de Pós-Graduação em Engenharia de Transportes:
Profª. Associada **Ana Paula Camargo Larocca**

Presidente da Comissão de Pós-Graduação:
Prof. Associado **Luis Fernando Costa Alberto**

# DEDICATÓRIA

*À minha família, os vizinhos imediatos mais perfeitos que eu poderia ter.*

# AGRADECIMENTOS

Minha eterna gratidão deve-se, certamente, ao meu melhor amigo, Deus, cuja presença constante em minha jornada supriu toda necessidade de força e confiança de que precisava, além da perseverança demandada sempre que o objetivo final pretendido era de difícil alcance. O Seu cuidado para comigo nos menores detalhes me inspirou a fazer tudo com dedicação e a Ele, definitivamente, sejam todas as congratulações provenientes desse trabalho.

Aos meus pais, Doralice e Hadenizio; irmãs, Sara, Débora e Daniela; e cunhado, Josué, cuja intercessão e torcida foram o combustível que me manteve em constante progresso no decorrer desses anos. Agradeço por me instigarem a sempre dar o meu melhor e me fortalecerem nos momentos de desânimo.

À minha orientadora Cira, pela amizade, confiança, apoio, companheirismo, bom-humor e por mostrar que sempre é possível melhorar. Ainda, grato pelo compartilhamento de experiências e ensinamentos que levamos para a vida.

Aos professores Bruno Bertoncini, Rita Salgueiro e Antonio Nélson, pelas valiosas contribuições ao presente trabalho.

Às minhas companheiras de orientação, pelos momentos agradáveis na sala de orientados, preocupação e solidariedade. Agradeço a Cintia, pela amizade e valiosas recomendações.

Aos meus professores da graduação, que criaram a base para que eu conseguisse ingressar no mestrado. Agradeço especialmente a Lílian, pelas parcerias e oportunidades que despertaram em mim o interesse pela pesquisa.

Aos meus amigos do STT, em particular os da turma 2017.1 e do Planejamento, pela parceria tanto dentro quanto fora da sala de aula. Sou grato aos queridos Juliana e Felipe, meus fiéis companheiros de açaí e cinema.

À turma V de Engenharia Civil da UFT e outros amigos de longa data da graduação, pela consideração e por torcerem tanto por mim.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo auxílio financeiro.

Às secretárias Ana e Sabrina, pela constante disposição para ajudar e contínua assistência.

Ao Renato Arbex, pelo contato com a SPTrans que permitiu a aquisição da pesquisa sobe/desce utilizada nesse trabalho. Agradeço, igualmente, à SPTrans, pela disponibilização dos dados.

A todos que, direta e/ou indiretamente, foram imprescindíveis para o desenvolvimento desse projeto bem como o amadurecimento acadêmico e pessoal do autor.

# EPÍGRAFE

"*If the ax is dull and its edge unsharpened, more strength is needed, but skill will bring success*."

*Ecclesiastes 10:10*

# RESUMO

MARQUES, S. F. **Estimativa do volume de passageiros ao longo de uma linha de transporte público por ônibus a partir da Geoestatística**. 2019. 98 p. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2019.

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 utiliza a dependência espacial entre as observações de viagens no intuito de gerar estimativas do volume de Embarques e Desembarques por ponto de parada, bem como Carregamento nos trechos, ao longo de uma linha de transporte público. Dados de uma pesquisa sobe/desce, realizada em oito linhas de ônibus da cidade de São Paulo, foram disponibilizados pela SPTrans, sendo que, selecionou-se a linha 856R-10 para compor uma análise do desempenho da Geoestatística, ferramental adequado para a estimativa de valores de variáveis em locais não amostrados, sobretudo em bancos de dados espacialmente dependentes. O processo de estimação foi realizado por meio da Krigagem Ordinária com distâncias euclidianas, técnica de interpolação geoestatística que, para sua aplicação, exige apenas o valor da variável em pontos do espaço e suas respectivas coordenadas geográficas. Posteriormente, tal método foi comparado às estimativas obtidas por meio da utilização de distâncias em rede ao longo da linha de ônibus. Os resultados obtidos, a partir da validação cruzada e métricas de aderência, apontaram um ajuste considerável dos valores observados e estimados para ambos os tipos de distância. Dessa forma, a investigação proposta confirmou, com sucesso, a viabilidade de aplicação da Geoestatística às variáveis de demanda por transporte público ao longo de uma linha de ônibus. Além disso, uma vez que a pesquisa sobe/desce exige recursos elevados para sua realização, os resultados destacam o potencial dos interpoladores geoestatísticos em eliminar a necessidade de execução do referido levantamento em todo o percurso da linha de ônibus, já que a Geoestatística garante estimativas aproximadas aos valores da pesquisa completa.



Palavras-chave: Estatística espacial. Pesquisa sobe/desce. Krigagem. Demanda por transporte público. Distâncias em rede. Distâncias euclidianas.

# ABSTRACT

MARQUES, S. F. **Estimation of passenger volume along a bus transit line using Geostatistics**. 2019. 98 p. Thesis (Master) – São Carlos School of Engineering, University of São Paulo, São Carlos, 2019.

The classical modeling of travel demand 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 relies on the spatial dependence between the observations of trips in order to generate estimates of the Boarding and Alighting passenger volume per bus stop and Loading in the stretches, along a public transport line, that would not be sampled for the survey occasion. Data from a passenger Boarding and Alighting survey carried out on eight bus lines in the city of São Paulo were made available by SPTrans, and line 856R-10 was selected to compose an analysis of the performance of Geostatistics, a tool that deals with the problem of scarce data and estimation in non-sampled places in the forecast of variables of public transport demand, especially in spatially dependent databases. The estimation process was performed by means of Ordinary Kriging with Euclidian distances, 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. Subsequently, this method was compared to the estimates obtained through the use of network distances along the bus line. The results obtained, from cross-validation and measures of goodness-of-fit, indicated a considerable adjustment of the observed values to the estimated ones for both types of distance. Thus, the proposed research successfully confirmed the feasibility of applying Geostatistics to the variables of public transport demand along a bus transit line. Moreover, since the boarding/alighting survey requires high resources for its realization, the results highlight the potential of the geostatistical interpolators in eliminating the need to carry out the survey in the whole course of the bus line, since Geostatistics provides estimates similar to the values of the complete survey.



Keywords: Spatial statistics. Boarding/alighting survey. Kriging. Public transport demand. Network distances. Euclidian distances.

# LISTA DE ILUSTRAÇÕES

Figura 1 – Mapa das linhas de ônibus contempladas pela pesquisa sobe/desce na cidade de São Paulo ..... 41
Figura 2 - Fluxograma das etapas que compõem o método do trabalho ..... 42
Figura 3 - Pesquisa de pares para cálculo do semivariograma empírico ..... 48
Figura 4 - Parâmetros do semivariograma a serem calibrados ..... 50
Figura 5 – Perfis de a) Embarques, b) Desembarques e c) Carregamento na linha 856R-10 ..... 57
Figura 6 – a) Autovalores mínimos para Embarques e Desembarques; e b) Autovalores mínimos para Carregamento ..... 60
Figura 7 – Porcentagem de desvio nas estimativas de krigagem para Embarques e Desembarques .... 63
Figura 8 - Porcentagem de desvio nas estimativas de krigagem para Carregamento ..... 66
Figura 9 - a) Semivariograma exponencial de Embarques; b) Semivariograma gaussiano de Embarques; c) Semivariograma gaussiano de Desembarques; d) Semivariograma gaussiano de Desembarques; e) Semivariograma esférico de Carregamento; e f) Semivariograma esférico de Carregamento ..... 67
Figura 10 – Desempenho de a) Embarques em linha reta; b) Embarques em rede; c) Desembarques em linha reta; d) Desembarques em rede; e) Carregamento em linha reta; e f) Carregamento em rede..... 70
Figura 11 –Semivariograma empírico de a) Embarques, b) Desembarques e c) Carregamento, os três calculados para o azimute 162° ..... 75
Figura 12 – Estimativa de a) e b) Embarques; e c) e d) Desembarques em pontos não amostrados .... 78
Figura 13 – Semivariogramas empíricos de Embarques (azimute entre parênteses na legenda do eixo das ordenadas) ..... 95
Figura 14 - Semivariogramas empíricos de Desembarques (azimute entre parênteses na legenda do eixo das ordenadas) ..... 96
Figura 15 - Semivariogramas empíricos de Carregamento (azimute entre parênteses na legenda do eixo das ordenadas) ..... 97
Figura 16 – Modelos teóricos para os semivariogramas de a) Embarques com distâncias euclidianas; b) Embarques com distâncias em rede; c) Desembarques com distâncias euclidianas; d) Desembarques com distâncias em rede; e) Carregamento com distâncias euclidianas; e f) Carregamento com distâncias em rede ..... 98

# LISTA DE TABELAS

Tabela 1 - Modelos típicos de semivariograma ..... 50
Tabela 2 - Métricas de aderência ..... 53
Tabela 3 - Recursos computacionais utilizados no estudo ..... 56
Tabela 4 - Medidas descritivas das variáveis de interesse ..... 58
Tabela 5 – Potências ótimas da transformação de Box-Cox ..... 59
Tabela 6 - Autocorrelação espacial das variáveis de interesse ..... 59
Tabela 7 - Desempenho de três modelos de semivariograma à estimativa das variáveis ..... 62
Tabela 8 – Testes de hipóteses não-paramétricos aplicados à comparação de resultados entre Krigagem Ordinária com distâncias euclidianas e com distâncias em rede ..... 72
Tabela 9 - Testes de hipóteses não-paramétricos aplicados à comparação de resultados entre Krigagem Ordinária omnidirecional e na direção de 162°, ambas com distâncias euclidianas ..... 77
Tabela 10 - Testes de hipóteses não-paramétricos aplicados à comparação entre os valores reais e as estimativas da Krigagem Ordinária omnidirecional com distâncias euclidianas ..... 93
Tabela 11 - Testes de hipóteses não-paramétricos aplicados à comparação entre os valores reais e as estimativas da Krigagem Ordinária omnidirecional com distâncias em rede ..... 94

SUMÁRIO

1. INTRODUÇÃO ..... 23
1.1. Objetivo geral ..... 25
1.2. Objetivos específicos ..... 25
1.3. Justificativa ..... 26
2. REVISÃO BIBLIOGRÁFICA ..... 29
2.1. Geoestatística ..... 29
2.2. Transporte público ..... 34
3. MATERIAIS E MÉTODO ..... 39
3.1. Base de dados ..... 39
3.2. Procedimento metodológico ..... 41
3.2.1 Tratamento da base de dados ..... 43
3.2.2 Análise exploratória tradicional e espacial ..... 44
3.2.3 Teoria das variáveis regionalizadas ..... 46
3.2.4 Cálculo do semivariograma empírico ..... 47
3.2.5 Ajuste do semivariograma teórico ..... 49
3.2.6 Validação cruzada ..... 52
3.2.7 Krigagem ..... 54
3.2.8 Recursos computacionais ..... 56
4 RESULTADOS E DISCUSSÕES ..... 57
4.1. Comparação entre estimativas provenientes de distâncias euclidianas e em rede ..... 61
4.2. Comparação entre estimativas provenientes de distâncias euclidianas com semivariogramas omnidirecionais e na direção principal ..... 74
4.3. Superfície de Krigagem ..... 78
5 CONCLUSÕES E RECOMENDAÇÕES PARA TRABALHOS FUTUROS ..... 81
REFERÊNCIAS ..... 87

APÊNDICE ..... 93

# 1. INTRODUÇÃO

O equilíbrio entre 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, usualmente depende 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 linhas 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, exige o conhecimento do perfil de movimentação de passageiros ao longo das linhas de ônibus (CEDER, 2007; FERRAZ; TORRES, 2004), informação que só pode ser obtida, com precisão, a partir de uma pesquisa sobe/desce.

Entretanto, levantamentos massivos, como a pesquisa sobe/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 com estimativas de dados espacialmente dependentes em locais não amostrados.

Tendo em vista os empecilhos à realização da pesquisa sobe/desce, as técnicas de interpolação espacial, empreendidas pela Geoestatística, podem permitir 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, seja 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.

Além disso, a teoria que dá suporte à Geoestatística aborda as variáveis espaciais em sua verdadeira natureza, ou seja, analisa o fenômeno regional como funções aleatórias e não determinísticas, como é o caso de técnicas de interpolação anteriores às geoestatísticas. Tal tratamento permite que a Geoestatística garanta resultados não enviesados e realize inferência estatística nas estimativas por meio do cálculo de variância.

Outra contribuição em potencial da aplicação da Geostatística à demanda por transporte público ao longo de linhas de ônibus se refere à incorporação da dependência espacial presente nessas variáveis à sua modelagem. Essa característica, que se manifesta quando observações próximas entre si no espaço possuem valores semelhantes, é comumente negligenciada nos modelos tradicionais não espaciais de demanda por transportes. Entretanto, o advento da Geostatística permitiu que a dependência espacial não somente fosse incluída no cálculo das estimativas como se tornasse a principal fonte de informações que dá suporte a esse ferramental.

Nesse âmbito, realizaram-se pesquisas que aplicam a Geoestatística a variáveis de demanda por transportes, as quais diferem, especialmente, pelo tipo de elemento agregador (suporte) utilizado na abordagem. Estudos que tratam a demanda por áreas (ROCHA; LINDNER; PITOMBO, 2017), domicílios (LINDNER; PITOMBO, 2017), trechos viários (SELBY; KOCKELMAN, 2013) e estações de transporte público por metrô (ZHANG; WANG, 2014) foram encontrados, porém, a cada um deles nota-se a ocorrência de restrições metodológicas, ora associadas à irregularidade dos suportes e/ou nível de agregação das variáveis, ora relacionadas à distribuição espacial dos dados para a modelagem ou ausência de aprofundamento na análise.

O tratamento geoestatístico da demanda por transporte público por ônibus, por sua vez, promete superar as limitações das abordagens anteriores. Os elementos agregadores podem ser considerados como regulares, as variáveis de interesse possuem um grau de agregação aceitável (nem tão elevado nem tão reduzido) e os pontos de coleta de dados para a realização da modelagem estão densamente dispostos no interior do espaço a ser interpolado. Além disso, atrelado às dificuldades inerentes ao levantamento do volume de Embarques, Desembarques e Carregamento ao longo de linhas de ônibus, o enquadramento dessa abordagem às suposições da Geoestatística sugere a obtenção de bons resultados. Sendo assim, a presente proposta apresenta contribuição metodológica, vinculada à aplicação da Geoestatística a dados de demanda por transporte público, e contribuição prática, relativa à possível economia de recursos da pesquisa sobre/desce.

Ademais, tendo em vista o fato de que os interpoladores espaciais utilizam o distanciamento entre pontos do banco de dados para o cálculo dos ponderadores associados a cada um dos pontos, a Geoestatística permite ainda duas opções diferentes de abordagem: (1) a de distâncias euclidianas, alternativa tradicional e (2) a de distâncias ao longo da rede de transporte. Uma vez que a atividade de transporte ocorre ao longo de uma rede viária, autores como Wang e Kockelman (2009) afirmaram que utilizar distâncias em rede em substituição às tradicionais euclidianas poderia melhorar os resultados dos interpoladores.

Porém, a aplicação de distâncias em rede não permite a inclusão da anisotropia do fenômeno na modelagem geoestatística. Para funções espaciais que variem de acordo com a direção considerada, desconsiderar essa característica no cálculo das estimativas pode afetar negativamente os resultados. Dessa forma, surgem dois questionamentos: para o caso das variáveis oriundas de uma pesquisa sobe/desce, a utilização de distâncias ao longo da linha de ônibus em detrimento das euclidianas melhora significativamente o desempenho do interpolador? E, se o fenômeno regionalizado é tratado como isotrópico, mesmo demonstrando ter uma direção principal bastante aparente, há alguma perda no poder preditivo da técnica?

## 1.1. Objetivo geral

A partir dos problemas discutidos anteriormente (restrições metodológicas de abordagens anteriores e custos elevados da pesquisa sobe/desce), o objetivo principal do presente estudo é estimar o volume de passageiros, ao longo de uma linha de transporte público por ônibus, a partir da Geoestatística.

Pretende-se, em outras palavras, encontrar respostas para o seguinte questionamento: é possível, por meio das técnicas de interpolação geoestatísticas, alcançar estimativas suficientemente precisas do número de Embarques e Desembarques por ponto de parada e Carregamento nos trechos de linha, de forma a tornar desnecessária a realização da pesquisa sobe/desce em toda a rede de TP?

## 1.2. Objetivos específicos

No percurso rumo à consecução do objetivo geral desse projeto, almeja-se também alcançar os seguintes objetivos específicos:

- Propiciar subsídios à não necessidade de realização da pesquisa sobe/desce em todo o itinerário das linhas de ônibus e/ou todas as linhas de determinada rede de TP;
- Comparar o desempenho de técnicas de interpolação espacial que utilizam distâncias euclidianas com aquelas que fazem uso de distâncias em rede;
- Verificar se há diferenças significativas no resultado das estimativas quando se ignora a direção principal em que o fenômeno regionalizado ocorre no espaço.

## 1.3. Justificativa

Aliado à importância da pesquisa de Embarque e Desembarque para o planejamento de redes de TP, o potencial ainda não explorado da aplicação da Geoestatística às variáveis de demanda ao longo de linhas de ônibus sugere a obtenção de bons resultados. Isso se deve ao fato de que, nesse caso, as unidades básicas de tratamento (pontos de parada e trechos de linha) obedecem aos pressupostos geoestatísticos (suportes homogêneos) e possuem um nível de agregação satisfatório, condição ainda não verificada em trabalhos anteriores (GOMES et al., 2016; LINDNER; PITOMBO, 2017; PITOMBO et al., 2015).

Ademais, é interessante notar que o padrão espacial do volume de passageiros que embarcam e desembarcam por ponto de parada não é um fenômeno isolado e totalmente aleatório. Antes, porém, esse comportamento está fortemente atrelado ao uso do solo no entorno das linhas de ônibus. Dessa forma, não somente assegurar redução de custos com levantamentos de campo, mas também conhecer de que forma as variáveis relativas ao ambiente construído afetam a geração de viagens, é de suma importância para a otimização do desempenho das redes de TP, pois subsidia tomadas de decisão quanto ao zoneamento e ordenamento territorial das cidades.

## 1.4. Estrutura do texto

Com base nos objetivos listados anteriormente, o Capítulo 2 busca trazer à tona o conhecimento dos conceitos básicos relacionados à Geoestatística e citar trabalhos que aplicaram esse ferramental para a estimativa/modelagem da demanda por transportes, observando os erros e acertos (vantagens e desvantagens) obtidos em cada exemplo no intuito de detectar eventuais lacunas de pesquisa existentes no estado da arte e/ou aspectos que ainda

podem ser melhorados. O Capítulo 3, por sua vez, 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, no Capítulo 4. As contribuições alcançadas pela solução proposta ao problema abordado são destacadas no Capítulo 5, no qual também são listadas sugestões para trabalhos futuros, a fim de fortalecer, consolidar e aprofundar as conclusões encontradas nesse trabalho.

# 2. REVISÃO BIBLIOGRÁFICA

Tendo em vista o escopo apresentado no capítulo anterior, esse capítulo se dedica a descrever as duas instâncias envolvidas no problema e solução propostos: transporte público e Geoestatística. Na primeira parte, estabelece-se uma evolução da aplicação da Geoestatística à estimativa/modelagem de variáveis de demanda por transportes, cujo objetivo é detectar falhas e/ou abordagens ausentes nos estudos encontrados até o presente momento. Posteriormente, a Seção 2.2 aponta as lacunas existentes na coleta de dados de demanda por transporte público, consolidando a justificativa que dá suporte à presente pesquisa.

## 2.1. Geoestatística

A Geostatística surgiu na década de 1950 como uma solução à estimativa de valores não amostrados na engenharia de minas (OLIVER; WEBSTER, 2015). Entretanto, sua versatilidade permitiu a ampliação da aplicação das técnicas geoestatísticas às áreas de epidemiologia (GOOVAERTS, 2009, 2012; HIBBERT et al., 2009), geologia (BURGESS; WEBSTER, 2006; LASLETT et al., 2006; OLIVER; WEBSTER, 2010), recursos hídricos (ZIMMERMAN et al., 1998), qualidade do ar (BAYRAKTAR; TURALIOGLU, 2005) entre outras. Na engenharia de transportes, o emprego dos conceitos geoestatísticos já foi verificado na modelagem de acidentes/segurança viária (GOMES et al., 2018; GUNDOGDU; FACULTY, 2014; MAJUMDAR; NOLAND; OCHIENG, 2004; MANEPALLI; BHAM, 2011) e, mais recentemente, na estimativa de variáveis de demanda por transportes (KLATKO et al., 2017; LINDNER; PITOMBO, 2017; SELBY; KOCKELMAN, 2013; ZHANG; WANG, 2014).

A utilização da Geoestatística para a estimativa de variáveis de demanda por transportes, para dados em rede, ainda se encontra bastante restrita à engenharia de tráfego, mais especificamente, à previsão do Volume Médio Diário Anual (VMDA) de trechos viários. Chi e Zheng (2013), por exemplo, interpolaram espacialmente, por meio da Krigagem Ordinária, a variável *transport footprint*, associada a cada trecho do sistema viário de Houghton County, Michigan, EUA. Os autores colocaram a variável de interesse em função do VMDA e, utilizando dados de postos de contagem tanto de Houghton quanto de três municípios adjacentes, obtiveram assim o modelo de variograma mais apropriado a partir de validação cruzada, pela técnica do ponto fictício. Os resultados não foram muito satisfatórios, uma vez

que a porcentagem de erro variava de acordo com a magnitude dos valores, isto é, os erros cresciam à medida que o valor real da variável de interesse aumentava.

Klatko et al. (2017) também se valeram da Krigagem Ordinária para interpolar dados de demanda por transportes. Esses autores aplicaram quatro diferentes métodos de interpolação espacial a fim de estimar o VMDA e, consequentemente, o total de milhas percorridas por veículos para vias locais sem dados de contagem de tráfego. Apesar de nenhuma das técnicas ter demonstrado um desempenho predominantemente superior ao das demais, os autores afirmam que a interpolação espacial pode levar a estimativas mais confiáveis e precisas do que outros métodos existentes.

Embora a Krigagem Ordinária demonstre um desempenho melhor que outras ferramentas de interpolação espacial, a suposição de que que a variável resposta possui média local constante em todo o espaço a ser interpolado pode comprometer o desempenho da técnica quando aplicada a bancos de dados que apresentam variações significativas (CHI; ZHENG, 2013). Para contornar esse problema, dispõe-se da Krigagem Universal, ferramenta que admite a existência de tendências no comportamento da variável de interesse.

Enquanto a Krigagem Ordinária exige apenas o valor das variáveis de interesse em pontos do espaço e as respectivas coordenadas geográficas, a Krigagem Universal refina esse estimador ao acrescentar informações adicionais à sua estrutura, também denominadas variáveis explicativas, cujo principal objetivo é aumentar o poder preditivo da técnica. Porém, ressalta-se que ambas as ferramentas dão origem a uma superfície contínua de valores estimados abrangendo, também, os pontos que não foram amostrados pela pesquisa.

Dessa forma, Eom et al. (2006) aplicaram a Krigagem Universal para a estimativa do VMDA de 5 tipos de vias de Wake County, EUA. Esses autores utilizaram dois métodos de calibração, WLS (*Weighted Least Squares*) e REML (*Restricted Maximum Likelihood*) e uma amostra com apenas 200 dos 1.154 postos de contagem de tráfego disponíveis. Os resultados dos dois modelos espaciais foram comparados aos de uma regressão linear múltipla (OLS, *Ordinary Least Squares*), sendo que a Krigagem Universal REML foi a técnica que apresentou os menores erros, seguida da WLS e da OLS. O desempenho superior da Krigagem Universal em relação às técnicas não espaciais também foi verificado por Subhadip et al. (2017).

A Krigagem Universal foi utilizada, ainda, por Wang e Kockelman (2009) para a estimativa do VMDA de duas diferentes classes de rodovias do estado do Texas, EUA. Os autores verificaram que o método aplicado retornava resultados razoavelmente confiáveis para volumes de tráfego intermediários, enquanto que, para baixos VMDA's e volumes elevados, a técnica superestimou e subestimou, respectivamente, os valores reais. Segundo eles, esses

resultados poderiam ser melhorados com a aquisição de dados acerca de variáveis explanatórias e utilização de distâncias em rede.

A inclusão de variáveis explicativas à Krigagem Universal para efeito de comparação com a Krigagem Ordinária foi estabelecida por Benedict, Eric e Joseph (2015), que se propuseram a testar o ajuste de três tipos de krigagem (simples, ordinária e universal) associados a três modelos de variograma (exponencial, esférico e gaussiano) aos dados de VMDA referentes aos anos de 2008, 2009 e 2010 do estado de Washington, EUA. No geral, baseado em um processo de validação no qual aplicaram-se várias métricas de erros, não foi identificada uma combinação ótima para o banco de dados. Dessa forma, os autores reforçaram que a melhor técnica de krigagem e melhor variograma só poderiam ser determinados a partir da estrutura presente nas informações disponíveis sobre a variável de interesse.

A utilização de distâncias em rede, por sua vez, foi abordada por Selby e Kockelman (2013) e Zhang e Wang (2014), sendo que, nesse âmbito, ambos os trabalhos alcançaram resultados convergentes. Os primeiros autores, além de incluírem distâncias em rede, ainda buscaram comparar o desempenho da Krigagem Universal ao de outra técnica espacial local, a Regressão Geograficamente Ponderada (GWR, *Geographically Weighted Regression*), na modelagem do VMDA para vias do estado do Texas nos Estados Unidos. Baseados em métricas de erros, aplicadas na etapa de validação, os resultados mostraram um desempenho ligeiramente melhor da Krigagem Universal em comparação à GWR, sendo que ambas as técnicas foram superiores a um terceiro modelo não espacial que considerava apenas a heterocedasticidade presente nos resíduos. Segundo os autores, a utilização de distâncias em rede, além de ter exigido expressivos recursos computacionais e de tempo de processamento, não demonstrou impacto positivo significativo nos resultados, embora eles afirmem que modelos baseados em dados de contagem, densamente distribuídos no espaço, poderiam ser beneficiados por tal recurso.

Já Zhang e Wang (2014) propuseram a aplicação da Krigagem Universal com distâncias euclidianas e em rede em alternativa à regressão linear tradicional para a estimativa do número médio de passageiros que entravam nas estações de metrô de Manhattan, por dia da semana, a fim de obterem a movimentação de passageiros em novas estações a serem implantadas posteriormente. Apesar de o $R^2$ ajustado não ter demonstrado grandes melhorias do modelo clássico para a krigagem com distâncias euclidianas e, finalmente, para a krigagem com distâncias em rede, as estatísticas *t* referentes aos parâmetros estimados na última técnica foram consideravelmente maiores, garantindo intervalos de confiança mais estreitos às estimativas provenientes da krigagem. Entretanto, corroborando os resultados de Selby e Kockelman

(2013), a pequena diferença no desempenho da krigagem com e sem distâncias em rede colocou, novamente, em dúvida a real eficácia da utilização de distâncias em rede.

Com base nas conclusões dos trabalhos citados acima, é possível afirmar que, além de a krigagem conseguir, realmente, alcançar excelentes resultados frente a modelos espaciais e não espaciais, sua aplicação à estimativa da demanda por transporte público pode atenuar os problemas identificados na previsão do VMDA. Em se tratando de uma rede de TP, os pontos de parada dos ônibus se distribuem espacialmente de forma muito mais regular e densa do que os postos de contagem de tráfego, fator que afeta sobremaneira o desempenho da técnica. Além disso, devido à escala menor das linhas de ônibus se comparada a rodovias federais e estaduais, o fenômeno regionalizado se mantém dentro dos limites do município, não tornando necessária a coleta de dados de municípios vizinhos, como ocorreu em um dos artigos mencionados (CHI; ZHENG, 2013). Ademais, tendo em vista o caráter mais "fechado" da rede de transporte público, ou seja, levando em consideração que, ao fim de uma viagem, o número de passageiros que embarcam será sempre igual ao volume de usuários que desembarcam, previne-se a existência de *outliers* no banco de dados, atenuando o problema de variância do erro que foi observado em alguns trabalhos (CHI; ZHENG, 2013; WANG; KOCKELMAN, 2009).

Outra potencial vantagem da aplicação dos conceitos geoestatísticos à movimentação de passageiros no transporte público por ponto de parada diz respeito ao nível de agregação da variável de interesse. Em se tratando de áreas urbanas, a demanda por transportes (ou outra variável relacionada) usualmente é trabalhada no âmbito de zonas de tráfego, variável agregada, ou domicílios, variável desagregada. Contudo, ambas as formas de abordagem podem interferir substancialmente na precisão e confiabilidade dos resultados. Quando a informação se encontra demasiadamente agregada, considera-se válida a afirmação de que o comportamento do fenômeno regional no âmbito mais agregado se repete de forma idêntica em cada elemento do nível mais desagregado, levando à ocorrência da falácia ecológica.

Dessa forma, as variações intrazonais demonstradas pela variável de interesse não conseguem ser capturadas pela modelagem geoestatística (LINDNER; PITOMBO, 2017). Lindner et al. (2016) também reconhecem que a atribuição de informações acerca de áreas a pontos (centroides) para a aplicação de técnicas geoestatísticas só é razoável quando os elementos agregados (nesse caso, as áreas) são pequenos em relação ao espaçamento da malha a ser interpolada, o que não se verifica em se tratando de zonas de tráfego.

Associado às barreiras identificadas anteriormente, existe ainda o problema da irregularidade no tamanho e forma das áreas normalmente utilizadas nas análises de demanda por transportes. Segundo Goovaerts (2008), a atribuição de informações referentes a elementos

desagregados (pontos) a unidades agregadas (áreas, representadas por seus respectivos centroides) supõe/assume, implicitamente, que tanto os elementos a serem utilizados para o cálculo do variograma quanto aqueles em que se pretende prever determinada variável possuem o mesmo tamanho e forma, condição que dificilmente é atendida pelas zonas de tráfego.

Por outro lado, no que concerne à demanda por transportes, a utilização de observações a um nível muito desagregado também não leva a resultados satisfatórios. Usualmente, quando se estuda viagens urbanas no âmbito de indivíduos ou domicílios, verifica-se a ocorrência de um elevado ruído no variograma, que se deve à alta variabilidade inerente ao comportamento humano. Estudos realizados por Lindner e Pitombo (2017), Gomes et al. (2016) e Pitombo et al. (2015), por exemplo, resultaram em uma variância, a distâncias muito pequenas, na preferência por determinado modo de transporte por domicílio que correspondia de 33% a 77% da variância máxima presente nessa variável. Em outras palavras, o valor da variável de interesse correspondente à uma parte substancial dos domicílios se diferenciava drasticamente de seus vizinhos imediatos. Segundo Landim (2006), variâncias aleatórias superiores a 30% na variável regionalizada são consideradas significativas, afetando fortemente a acurácia das estimativas geoestatísticas associadas.

Em contrapartida, os pontos de parada de ônibus coletivo, além de oferecerem um suporte aproximadamente regular, também permitem um grau de agregação nem tão elevado quanto o das áreas (zonas de tráfego), nem exageradamente pequeno como o de indivíduos ou domicílios. A mesma afirmação também pode ser estendida aos trechos das linhas de ônibus compreendidos entre dois pontos de parada consecutivos. Além disso, o estudo da movimentação de passageiros no transporte público que tem como base os dois suportes mencionados anteriormente (pontos de parada e trechos de linhas de ônibus) permite a aplicação da geoestatística em rede, ou seja, diretamente aos elementos de interesse do planejamento de TP. Esse assunto é discutido mais detalhadamente na seção a seguir.

Finalmente, levando-se em conta os trabalhos mencionados anteriormente, pode-se 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 associada a tais suportes; (2) estimar variáveis de redes de transporte urbanas e não intermunicipais (como VDMA – trabalhos anteriores) ; (3) realizar comparações das estimativas realizadas a partir de distâncias em redes e euclidianas.

## 2.2. Transporte público

De acordo com Rodrigue, Comtois e Slack (2017), a atividade de transporte é composta por dois atores, a saber: operadores, ou seja, fornecedores do serviço, e por aqueles que utilizam tal serviço, os usuários. Dada a importância do transporte para a economia e qualidade de vida (MEYER, 2016), o funcionamento adequado da sociedade exige que oferta e demanda estejam em equilíbrio. Esse equilíbrio só pode ser alcançado a partir do processo de planejamento de transportes.

No que concerne ao transporte público no Brasil, é notável que o planejamento e, consequentemente, desempenho dessa rede possui estreita relação com a mobilidade das camadas populacionais mais vulneráveis, o que faz do TP um instrumento de integração social e igualdade no uso do espaço público e na acessibilidade. A rede de TP, além do usuário e operador, possui ainda outra instância que também está envolvida na concessão do serviço: o poder público.

Dessa forma, no intuito de beneficiar, simultaneamente, usuário, poder público e operador, o planejamento de uma rede de transporte público engloba três principais categorias de objetivos, a saber: otimizar o desempenho do serviço de transporte, o que implica a provisão de altas velocidades nas viagens, conveniência ao usuário e outras facilidades que atraem passageiros; alcançar máxima eficiência operacional, ou seja, custos mínimos para determinado nível de serviço; e criar impactos positivos tais como: redução de congestionamentos, elevada mobilidade, padrões desejáveis de uso do solo, sustentabilidade, qualidade de vida, etc. (VUCHIC, 2005).

Contudo, tais alvos só podem ser atingidos a partir da elaboração de um planejamento essencialmente pautado no diagnóstico da rede de transporte público em consideração. Essa avaliação, por sua vez, só pode ser concebida por meio da realização de pesquisas e levantamentos nas linhas de ônibus coletivo, pontos de parada, estações, domicílios entre outros elementos envolvidos no serviço de TP. A pesquisa de origem-destino das viagens e dos desejos de viagens, pesquisa de variação da demanda ao longo do dia, pesquisa da confiabilidade e pontualidade, pesquisa de opinião dos usuários (FERRAZ; TORRES, 2004), pesquisa de velocidade e retardamento e pesquisa de Embarque e Desembarque (VUCHIC, 2005) são exemplos de investigações cujo objetivo é justamente prover subsídios à criação de políticas públicas direcionadas ao sistema de transporte público.

Dentre as pesquisas, mencionadas anteriormente, 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. Entretanto, levantamentos massivos, como a pesquisa sobe/desce e de origem-destino, demandam tempo, planejamento e investimentos financeiros, o que inclui a contratação de uma equipe de profissionais qualificados para a visitação das linhas de ônibus.

A fim de conceder uma solução alternativa à pesquisa sobe/desce, os contadores automáticos de passageiros (CAP's) surgem como uma alternativa promissora. Segundo Barbosa et al. (2006), Estados Unidos e Canadá são exemplos de países que já fazem uso dessa tecnologia, porém, em se tratando de países em desenvolvimento e outros menos evoluídos, a aplicação e consolidação do CAP para a contagem automática de passageiros encontram várias barreiras, o que leva a pesquisa sobe/desce a ser ainda o método mais amplamente utilizado para a coleta de informações a respeito da movimentação de passageiros ao longo de linhas de ônibus.

Barbosa et al. (2006) apresentaram e validaram um protótipo de CAP para o Brasil, baseado em placas metálicas, que seriam instalados nas portas dos veículos de transporte público. Documentos referentes à São Paulo Transporte S.A. trazem indícios da aquisição de tais aparelhos para a sua implantação nas linhas de ônibus da capital paulista, contudo, a custos elevadíssimos (SPTRANS, 2015). Tendo em vista as restrições de orçamento geralmente presentes nas cidades menores e de porte médio, para essas, a pesquisa sobe/desce permanece como único e principal recurso à coleta de dados para o diagnóstico e planejamento da rede de TP.

Uma segunda estratégia, menos onerosa, seria a obtenção/derivação das informações usualmente coletadas em uma pesquisa sobe/desce a partir da associação entre dados da bilhetagem eletrônica com os 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; WHITE, 2005). Mais uma vez, a pesquisa sobe/desce não pode ser totalmente substituída.

Em suma, a pesquisa sobe/desce requer a permanência de dois a três pesquisadores dentro de uma linha de ônibus contando e anotando o número de passageiros que embarcam e desembarcam em cada ponto de parada, os quais devem possuir um código específico que permita sua associação ao volume de Embarques e Desembarques coletado pelos

pesquisadores. Tendo em vista os elevados custos de uma pesquisa sobe/desce, a proposta empreendida nesse trabalho converge com o objetivo principal da Geoestatística, a saber: retirar o máximo de informação possível acerca de uma variável e, a partir da análise espacial, estimar seus valores para pontos não amostrados. Ou seja, o presente trabalho pretende 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 dos trechos), supondo que, para uma determinada rede de TP, nem todos os pontos, trechos e/ou linhas de ônibus foram amostrados. Esta seria a quarta lacuna de pesquisa associada ao presente trabalho.

Recapitulando, quatro abordagens da Geoestatística aplicada a variáveis de demanda por transportes foram encontradas, que podem ser classificadas de acordo com o elemento agregador utilizado na modelagem, a saber: áreas, domicílios ou indivíduos, trechos viários e estações de transporte público por metrô. A primeira, que utiliza áreas como suporte, possui restrições metodológicas associadas ao seu alto grau de agregação e à irregularidade no tamanho e forma das unidades de tratamento, as zonas de tráfego. A segunda abordagem supera as limitações da modelagem anterior, mas sofre com o nível de desagregação muito elevado dos domicílios e indivíduos, cujo alta aleatoriedade atrelada ao comportamento humano impede a obtenção de resultados totalmente satisfatórios pelos estimadores geoestatísticos.

A abordagem com trechos viários consegue contornar os problemas dos dois primeiros tratamentos, porém, surge uma nova limitação. A distribuição espacial não uniforme dos postos de contagem de tráfego gera, nas estimativas, uma variância dos erros, que serão menores apenas nas regiões com uma alta densidade de postos. As estações de transporte público, por sua vez, por estarem separadas por uma distância aproximadamente igual, levariam a resultados potencialmente melhores que os das modelagens precedentes. Entretanto, o único estudo, encontrado até o presente, que realiza essa abordagem, trabalhou apenas com a variável de Embarques. Além disso, 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.

Dessa forma, percebeu-se que não há, ainda, estudos isentos das restrições identificadas nas quatro abordagens anteriores e que investigue, de fato, o potencial da Geoestatística em estimar variáveis de demanda por transportes que possuam limitação de acesso, ou seja, cujos dados possam ser considerados como escassos. Essas lacunas convergem com a proposta empreendida no presente trabalho, uma vez que, no caso do transporte público por ônibus, o

perfil de movimentação de passageiros pode ser levantado, com precisão, apenas por meio da pesquisa sobe/desce, cuja realização demanda elevados custos para a municipalidade. Além disso, os elementos de tratamento, nesse caso (pontos de parada e trechos de linha), podem ser considerados como regulares, e as variáveis de interesse a eles associadas (Embarques, Desembarques e Carregamento) não estão exageradamente agregadas, como no caso das áreas, nem tão desagregadas, como na abordagem de domicílios e indivíduos.

Em síntese, as seguintes lacunas de pesquisa foram enumeradas: (1) estimar variáveis de demanda por transportes, a partir da Geoestatística, com suportes homogêneos (pontos de parada, por exemplo) e baixa variabilidade associada a tais suportes; (2) estimar variáveis de uma rede de transporte urbana e não intermunicipal (como VDMA – trabalhos anteriores); (3) realizar comparações das estimativas realizadas a partir de distâncias em redes e euclidianas; e (4) 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 dos trechos), supondo que, para uma determinada rede de TP, nem todos os pontos, trechos e/ou linhas de ônibus foram amostrados. O Capítulo 3 introduz os materiais e métodos que consolidam a solução proposta às lacunas citadas.

# 3. 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.

## 3.1. Base de dados

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 sobe/desce, na qual sejam 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 em 2012 (abrangendo 631 pontos de parada) 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 da SPTrans 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

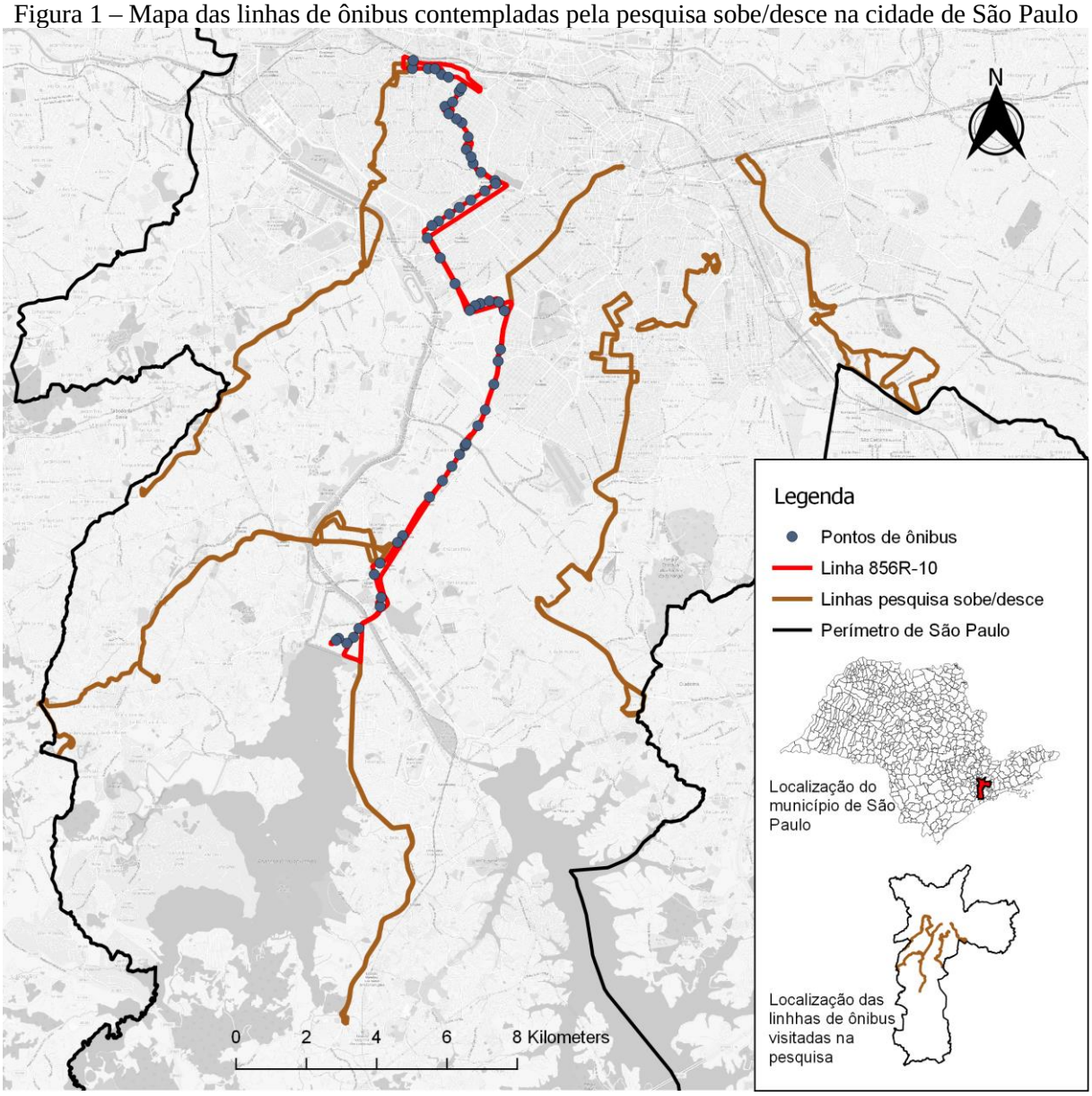
como: código identificador da linha e da viagem e um número associado ao sentido da viagem, ou seja, ida ou volta.

A fim de promover uma avaliação do potencial das ferramentas geoestatísticas na resolução do problema de escassez de dados de volume de passageiros ao longo de linha de ônibus, foi selecionada, dentre as oito linhas visitadas por ocasião da pesquisa sobe/desce em São Paulo, apenas uma linha. O comportamento espacial das oito linhas, tanto em relação ao número de usuários que subiam quanto que desciam em cada ponto, bem como ao Carregamento dos trechos entre pontos de parada consecutivos, foi discriminado em uma análise visual.

Mapas temáticos de Embarques e Desembarques, por ponto de parada, e Carregamento nos segmentos de linha foram gerados no *software* ArcGIS 10.1 para as oito linhas no sentido de ida e de volta e na faixa horária das 05:00 às 08:59, que compreende o horário de pico da manhã. A inspeção visual buscou identificar a combinação de linha e sentido cujos mapas apontaram uma estrutura espacial aparente, ou seja, variações suaves no valor das variáveis em sua vizinhança espacial, de forma que observações situadas próximas entre si apresentassem valores semelhantes. Um exemplo de estrutura espacial aparente seria o caso tradicional em que o Carregamento demonstra pequenos valores nos primeiros trechos da linha, aumenta até alcançar seu volume máximo em algum segmento intermediário e volta a diminuir conforme se aproxima o fim do percurso.

Nesse âmbito, a linha 856R-10 foi escolhida por apresentar um perfil de movimentação de passageiros que poderia gerar resultados satisfatórios 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, 57 (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.

Figura 1 – Mapa das linhas de ônibus contempladas pela pesquisa sobe/desce na cidade de São Paulo




Fonte: Elaboração própria a partir de dados do GeoSampa, IBGE e *Open Street Map*

## 3.2. Procedimento metodológico

O fluxograma, mostrado na Figura 2, sintetiza a sequência de etapas que compõem o método desse trabalho. Em suma, três análises principais foram realizadas:

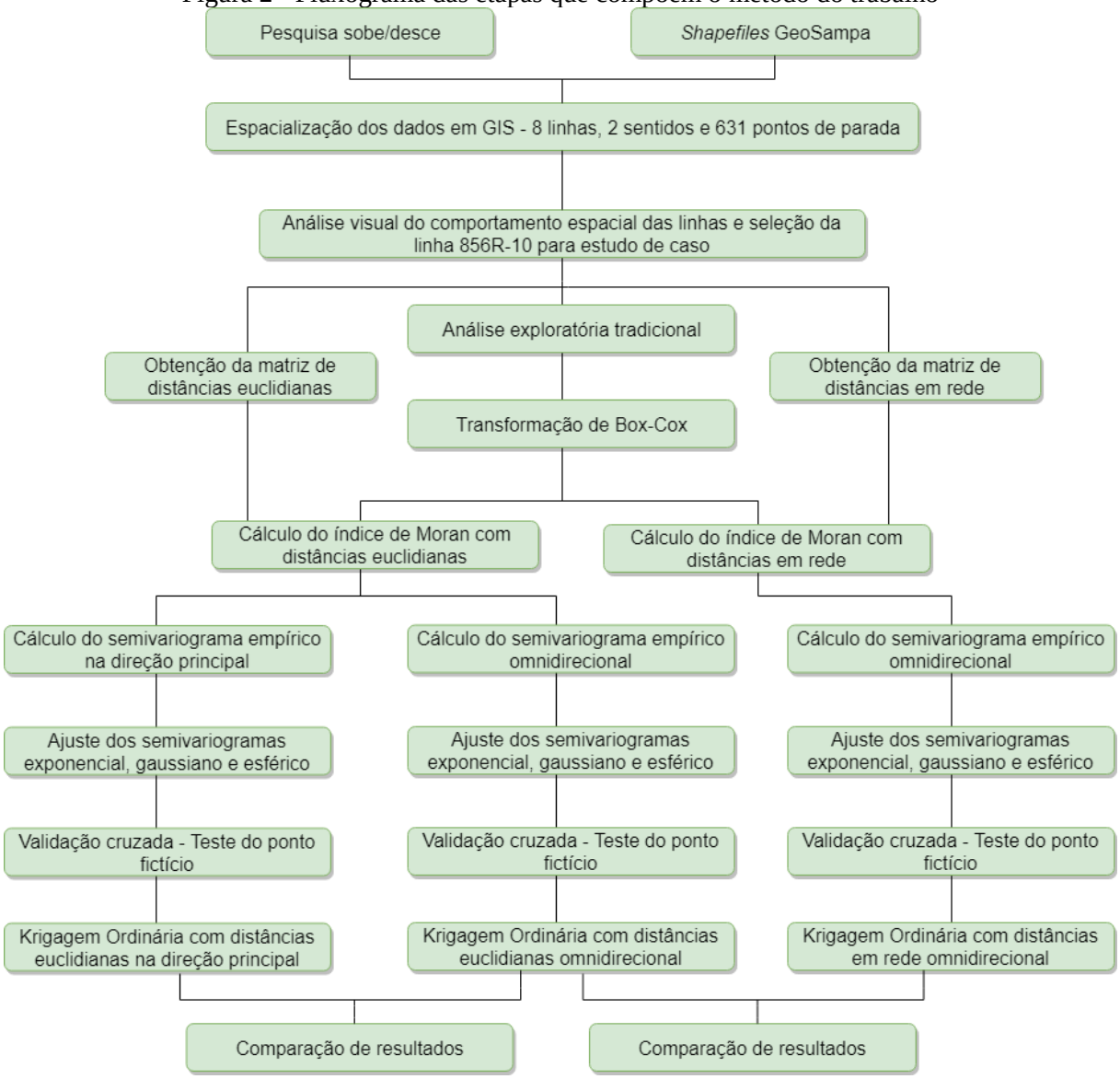
(1) Estimativas de variáveis de demanda por TP a partir de Krigagem Ordinária e semivariograma omnidirecional;

(2) Comparação entre os resultados da Krigagem Ordinária com distâncias euclidianas e em rede; e

(3) Comparação entre os resultados provenientes da Krigagem Ordinária (com distâncias euclidianas) e semivariograma omnidirecional e os provenientes do semivariograma na direção principal.

Para tanto, algumas etapas se antecederam, tal como a verificação de assimetria e posterior correção desta a partir da transformação de Box-Cox e a avaliação de autocorrelação espacial das variáveis de interesse por meio do índice de Moran. Todas essas fases estão detalhadamente descritas nas subseções seguintes.

Figura 2 - Fluxograma das etapas que compõem o método do trabalho




Fonte: Elaboração própria

### 3.2.1 Tratamento da base de dados

A espacialização do banco de dados foi realizada por meio do Sistema de Consulta do Mapa Digital da Cidade de São Paulo, hospedado no sítio eletrônico do GeoSampa. Dessa forma, teve-se acesso às malhas digitais (arquivo *shapefile*) das linhas de ônibus e de pontos de parada de São Paulo. De posse dos códigos identificadores das oito linhas de ônibus, contempladas pela pesquisa sobe/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 das coordenadas geográficas dos pontos de parada e pontos médios dos trechos e do itinerário da linha 856R-10 foi possível, posteriormente, gerar as matrizes de distância euclidiana e distância em rede entre todos os pontos. As matrizes de distâncias euclidianas e de distâncias em rede foram obtidas, respectivamente, por meio dos pacotes computacionais abertos e livres QGIS 2.18.12 e *Geographic Resources Analysis Support System* (GRASS GIS 7.4.0). Para as variáveis Embarques e Desembarques, as matrizes tiveram dimensão de 57x57, ao passo que, para Carregamento, teve-se uma ordem de 56, já que o segmento de linha a jusante do último ponto de parada pertence ao percurso de volta da linha. No total, foram produzidas quatro matrizes.

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; e distância em linha reta e ao longo da linha 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 Seção 3.2.2.

### 3.2.2 Análise exploratória tradicional e espacial

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 de tendência central e dispersão. No que concerne à demanda por transportes, usualmente trabalhada/fornecida na forma de contagens, nota-se, em geral, a ocorrência de uma assimetria positiva bastante proeminente, ou seja, fuga da normalidade.

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

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

em que $\dot{y} = \ln^{-1}\left[(1/n)\sum \ln y\right]$ 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 $\lambda$ pode ser encontrado a partir da otimização de alguma função objetivo relacionada à distribuição gaussiana. Essa função pode ser a verossimilhança, a estatística do teste de normalidade de Shapiro-Wilk (SHAPIRO; WILK, 1965) ou até mesmo o coeficiente de correlação entre as observações da variável transformada e a reta da distribuição normal no gráfico quantil-quantil (VOGEL, 1986), alternativa selecionada para ser aplicada às variáveis de demanda a serem trabalhadas.

Cabe ressaltar, entretanto, que como para um valor de $\lambda$ nulo, a transformação de Box-Cox equivale à utilização do logaritmo natural dos dados brutos, foi necessário somar uma unidade às variáveis número 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.

Após essa operação, as variáveis resultantes da transformação de Box-Cox passaram por uma etapa de análise exploratória de dependência espacial, que poderia fornecer indícios da eventual 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, já que se trata de um indicador que não exige recursos computacionais elevados e é simples de entender. Outra vantagem desse índice, cuja fórmula é mostrada na Equação 2, é que ele permite a avaliação de sua significância por meio do cálculo de um pseudo valor *p*, de interpretação semelhante ao dos testes estatísticos convencionais.

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

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

$w_{ij}$: ponderador que reflete a intensidade de interação espacial entre os pontos $i$ e $j$. Utilizou-se, no presente trabalho, o inverso da distância euclidiana e o inverso da distância ao longo da linha.

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 condição em que todos os pontos contidos no banco de dados apresentam 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 do ferramental geoestatístico, cuja teoria básica é explicada na seção seguinte.

### 3.2.3 Teoria das variáveis regionalizadas

Os primeiros trabalhos que abordam o problema da estimativa de variáveis em locais não amostrados datam dos anos 50 e são especialmente dedicados à engenharia de minas (KRIGE, 1951). Porém, o formalismo matemático da teoria das variáveis regionalizadas foi consolidado apenas na década seguinte pelo professor francês George Matheron (1963, 1971) ao notar que os métodos tradicionais de interpolação falhavam em considerar a variabilidade e dispersão, bem como o aspecto espacial das distribuições de probabilidade das variáveis relacionadas à mineração.

Segundo Matheron (1963), uma variável regionalizada, que não se resume apenas às características de um corpo de minério, mas pode ser encontrada em diversos outros campos de estudo, refere-se a uma função à qual atribui-se um valor definido em cada ponto do espaço. Dessa forma, uma variável regionalizada deve ser localizada, pode apresentar uma continuidade em sua variação espacial razoavelmente estável e possuir diferentes formas de anisotropias, ou seja, é capaz de se alterar de diversas maneiras conforme a direção considerada. Além disso, Matheron (1971) ainda ressalta duas características contraditórias, mas complementares das variáveis regionalizadas: elas demonstram um aspecto aleatório, que se destaca pela irregularidade e variações não previstas de um ponto para outro; e uma parte estruturada, que reflete o comportamento sistematizado do fenômeno espacial.

Ao formular a teoria das variáveis regionalizadas, Matheron (1971) buscava não somente descrever a correlação espacial de tais funções, mas também estabelecer a geoestatística, ferramental definido por ele mesmo como a aplicação da teoria das variáveis regionalizadas à estimativa de valores em locais não visitados a partir de dados amostrais fracionados. Por conseguinte, variáveis regionalizadas não relacionadas à mineração, como a demanda por transportes, por exemplo, puderam também tirar proveito dos conceitos e formulações introduzidas pela teoria supracitada.

Contudo, variáveis localizadas no espaço apresentam uma peculiaridade que exige, por parte da formalidade estatística, a adoção de uma importante hipótese para a Geoestatística. Trata-se do fato de que o valor da variável regionalizada correspondente aos pontos amostrados consiste em apenas uma observação da variável aleatória em consideração, já que a cada ponto no espaço está associada uma variável aleatória. Sendo assim, torna-se impossível avaliar a dispersão (por meio do cálculo de variância) e realizar inferência estatística somente com tais dados coletados. Dessa forma, além de se tornar necessário adotar que todos os pontos no

espaço possuem a mesma média, assume-se que essa medida de tendência central é constante, ou seja, que as observações amostradas, que seriam variáveis aleatórias individuais/particulares, constituem dados de uma única variável aleatória. Essa hipótese, conhecida como estacionariedade de segunda ordem (MATHERON, 1971; CRESSIE, 1993; JOURNEL; HUIJBREGTS, 1978), estabelece que a variância presente nas variáveis aleatórias espaciais não mais depende da posição geográfica dos pontos, mas sim da distância entre eles, possibilitando, assim, a execução de inferência estatística, uma das principais contribuições idealizadas pela teoria das variáveis regionalizadas.

### 3.2.4 Cálculo do semivariograma empírico

O semivariograma $\gamma(h)$, ou variograma $2\gamma(h)$, é a principal ferramenta gráfica da geoestatística que permite visualizar as características qualitativas da regionalização. Em outras palavras, permite conhecer a forma com que a variável regionalizada sob consideração se comporta espacialmente. A partir da hipótese de estacionariedade de segunda ordem, a função semivariograma deixa de depender da localização geográfica dos pontos amostrais e passa a se relacionar apenas à distância entre dois pontos contidos no banco de dados. Dessa forma, o cálculo do semivariograma empírico, ou experimental, é dado pela seguinte equação:

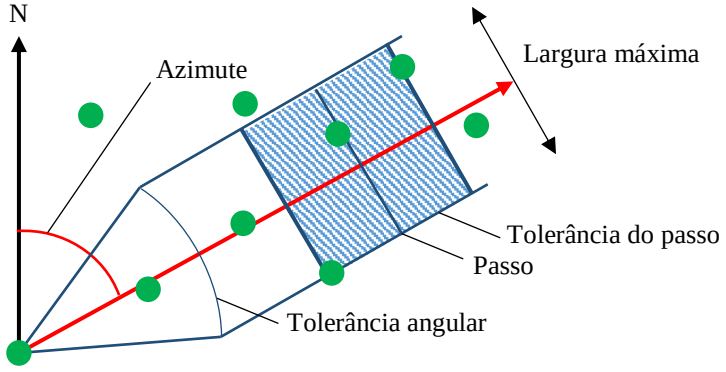
$$\gamma(h) = \frac{1}{2N} \sum_{i=1}^{N} \left[ Z(x_i + h) - Z(x_i) \right]^2 \tag{3}$$

em que *Z(x)*: valor da variável regionalizada *Z* na posição geográfica *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 (área hachurada na Figura 3) dentro da qual busca-se encontrar eventuais pontos a comporem, juntamente com o ponto de referência, os pares sobre os quais o semivariograma será calculado (JOURNEL; HUIJBREGTS, 1978; OLIVER; WEBSTER, 2015; YAMAMOTO; LANDIM, 2015). A Figura 3 ilustra o procedimento.

Figura 3 - Pesquisa de pares para cálculo do semivariograma empírico




Fonte: Yamamoto e Landim (2015)

No presente trabalho, utilizou-se, incialmente, 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 e, consequentemente, 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. Além disso, é possível afirmar que, nesse caso, o semivariograma omnidirecional traz respostas semelhantes às do semivariograma que considera a direção aproximada da linha. Ademais, o semivariograma omnidirecional consegue abranger uma maior quantidade de pares, o que pode melhorar a precisão dos resultados.

A escolha inicial do semivariograma omnidirecional se deu, também, em função da necessidade de uniformização do método para o cálculo das estimativas tanto para as distâncias euclidianas quanto para aquelas ao longo da linha de ônibus. Haja vista que, para o caso das distâncias em rede, os códigos computacionais se tornam mais complexos (não há, ainda, bibliotecas computacionais para esse fim) quando se adota uma direção específica, considerou-se o fenômeno como isotrópico a fim de tornar comparáveis os resultados obtidos a partir das distâncias euclidianas com os provenientes das distâncias em rede.

Entretanto, no intuito de verificar se as estimativas se diferenciam significativamente entre o caso em que se utiliza o semivariograma omnidirecional com aquele em que o fenômeno é analisado na direção principal, a interpolação espacial com distâncias euclidianas também foi realizada considerando a direção de maior continuidade espacial da regionalização. Dessa forma, tais resultados puderam ser comparados aos do caso isotrópico, ou seja, omnidirecional.

A identificação do azimute que caracteriza a direção principal do fenômeno, também conhecida como a orientação na qual a autocorrelação espacial atua continuamente por uma distância maior que para as outras direções, foi alcançada por meio de um processo exploratório. Calcularam-se, com o auxílio da linguagem de programação R (R CORE TEAM, 2018), os semivariogramas experimentais para as três variáveis de interesse variando-se o azimute da direção em intervalos de 18º graus de forma a se obter 10 gráficos para cada variável, iniciando-se no azimute 18º e finalizando na direção 180º. Essa amplitude de variação é suficiente para contemplar todas as orientações possíveis, haja vista que os ângulos situados no primeiro quadrante abrangem, também, os do terceiro, ao passo que as direções do segundo quadrante englobam as do quarto.

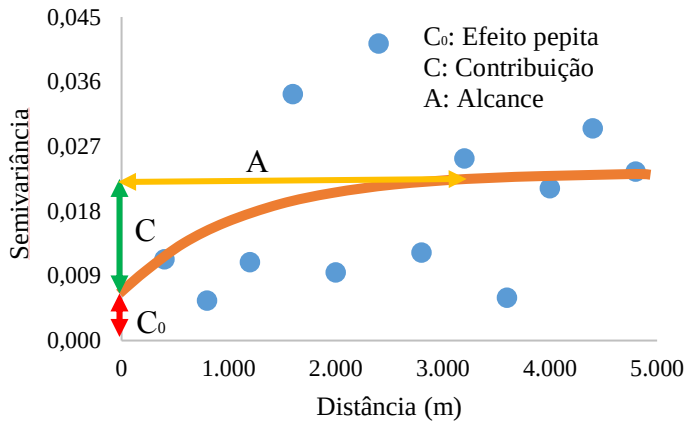
Os semivariogramas resultantes passaram por uma inspeção visual cujo objetivo era encontrar o gráfico com maior número de pontos, maior alcance e visível aproximação dos pontos de uma curva imaginária que representaria o comportamento da variável regionalizada, ou seja, semivariâncias crescentes a partir da origem até chegar a um limite, o patamar. Dessa forma, o azimute selecionado foi aquele que conseguia cumprir, simultaneamente e para as três variáveis de interesse, os três requisitos mencionados anteriormente.

No intuito de otimizar o desempenho do estimador a ser aplicado, mantiveram-se todos os pontos da linha 856R-10 para o cálculo do semivariograma empírico. Após esse cálculo, ajusta-se um modelo teórico aos valores resultantes do semivariograma experimental, tema a ser discutido na próxima seção.

### 3.2.5 Ajuste do semivariograma teórico

O processo de ajuste de uma função bem definida aos pontos do semivariograma experimental consiste na obtenção de três principais parâmetros a partir de determinado método pré-estabelecido. Tais parâmetros, efeito pepita, contribuição e alcance, são apontados na Figura 4.

Figura 4 - Parâmetros do semivariograma a serem calibrados




Fonte: Adaptada de Matheron (1963)

O efeito pepita ($C_0$) refere-se ao valor da semivariância quando a distância entre dois pontos tende a 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 o patamar parcial, ou contribuição (C), 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. Uma vez que, a partir dessa distância, a variância alcança seu valor máximo, assume-se que não há mais influência dos pontos entre si. Os três principais modelos teóricos de semivariograma são mostrados na Tabela 1.

Tabela 1 - Modelos típicos de semivariograma

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

Fonte: Adaptada de Olea (2006)
Nota: $C_0$, $C$ e $a$ equivalem, respectivamente, ao efeito pepita, patamar parcial e alcance.

A aproximação do semivariograma empírico às equações acima pode ser dada de diversas formas: verossimilhança, verossimilhança restrita (REML, *Restricted Maximum Likelihood*), mínimos quadrados (OLS, *Ordinary Least Squares*), mínimos quadrados ponderados (WLS, *Weighted Least Squares*), mínimos quadrados generalizados (GLS, *Generalized Least Squares*) entre outros (CRESSIE, 1993), além do ajuste visual. Enquanto os dois primeiros métodos repousam fortemente na suposição de normalidade para a variável de interesse, o terceiro método é puramente numérico, não reconhecendo a variação e covariação da distribuição de $\gamma(h)$. O critério de GLS, por sua vez, busca contabilizar o efeito da heterocedasticidade, ou seja, da ocorrência de variância não constante nos resíduos do modelo ajustado.

O ajuste do semivariograma empírico a um modelo teórico por meio da minimização dos mínimos quadrados ponderados foi introduzido por Cressie (1985) como uma técnica alternativa situada entre os critérios OLS e GLS, ou seja, nem tão grosseira quanto a primeira nem tão robusta e dispendiosa computacionalmente quanto a segunda. Segundo Cressie (1985), a técnica de WLS, que foi utilizada no presente trabalho, permite resultados satisfatórios, uma vez que se trata de um estimador estatisticamente sensível dos parâmetros do semivariograma teórico (efeito pepita, contribuição e alcance), podendo servir, também, como uma estimativa inicial ao processo iterativo de estimação pelo método generalizado. Além disso, em se tratando dos mínimos quadrados ponderados, as semivariâncias médias calculadas por meio da Equação 3 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 que o OLS.

Em suma, os semivariogramas experimentais de Embarques, Desembarques e Carregamento foram ajustados aos modelos exponencial, gaussiano e esférico tanto para abscissas com distâncias euclidianas quanto para distâncias ao longo da linha de ônibus. 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. Realizou-se, também, a calibração dos parâmetros efeito pepita, patamar parcial e alcance para os semivariogramas empíricos calculados considerando-se a direção identificada na inspeção visual (Seção 3.2.4), sendo que a adequação dos pontos às curvas listadas na Tabela 1 também foi utilizada como critério para a seleção desse azimute principal.

No que concerne à utilização de distâncias em rede na modelagem geoestatística, cabe ressaltar que alguns modelos teóricos podem se caracterizar como não permissivos, designação atribuída a eles quando as curvas dão origem a estimativas com variâncias negativas. Tal situação depende, basicamente, do arranjo espacial do conjunto de pontos a serem utilizados e dos segmentos de linha que os conectam, sobre os quais as distâncias serão calculadas. Ver Hoef (2018) elaborou códigos, também na linguagem de programação R, para verificar a ocorrência desse inconveniente de acordo com a curva selecionada para o semivariograma experimental.

Para tanto, avalia-se o comportamento dos autovalores da matriz de semivariância da krigagem conforme variações no parâmetro alfa, ou alcance, do modelo. Os autovalores podem ser definidos como números que, subtraídos da diagonal principal de uma matriz, a transforma em uma matriz singular, isto é, não invertível. A ausência de variância negativa nas estimativas dar-se-á somente se, para qualquer que seja alfa, não houver autovalor negativo. A partir dos códigos de Ver Hoef (2018), realizou a análise de autovalores para as distâncias em rede de Embarques e Desembarques, por meio da matriz de dimensão 57, e de Carregamento, utilizando-se a matriz de ordem 56.

A otimização dos parâmetros do semivariograma teórico também pode ser empreendida por meio da aplicação de métricas de erros aos resultados obtidos pelo estimador geoestatístico. Esse recurso, que se vale da comparação entre os valores reais e previstos pela modelagem, também é conhecido como validação cruzada e será melhor detalhado na Seção 3.2.6

### 3.2.6 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 calcular 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) | $\sum_{i=1}^{N}\left[y(s_i)-y^*_{-i}(s_i)\right]^2$ | (7) | - |
| Erro Médio (ME) | $\frac{1}{n}\sum_{i=1}^{n}\left[y^*_{-i}(s_i)-y(s_i)\right]$ | (8) | Indica a existência de erros sistemáticos |
| Erro Médio Absoluto (MAE) | $\frac{1}{n}\sum_{i=1}^{n}\left\lvert y(s_i)-y^*_{-i}(s_i)\right\rvert$ | (9) | Não particularmente sensitivo a grandes erros |
| Raiz Quadrada do Erro Quadrado Médio (RMSE) | $\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[y(s_i)-y^*_{-i}(s_i)\right]^2}$ | (10) | Grandes erros são duramente penalizados |
| Coeficiente de correlação linear de *Pearson* (R) | $\frac{1}{n}\sum_{i=1}^{n}\frac{\left[y(s_i)-\bar{y}\right]\left[y^*_{-i}(s_i)-\bar{y}^*\right]}{\sigma_y\sigma_{y^*}}$ | (11) | - |

Fonte: Adaptada de Hollander e Liu (2008)
Nota: $y(s_i)$ e $y^*_{-i}(s_i)$ representam os valores observados e estimados, respectivamente. O subscrito *-i* alerta que o valor real da variável no ponto *i* não é utilizado no cálculo da estimativa em *i*.

Cabe relembrar que, no total, foram obtidas nove estimativas para Embarques e Desembarques por ponto de parada e Carregamento no ponto médio dos trechos a partir do teste do ponto fictício. São elas:

(1) Estimativas com distâncias euclidianas para os três modelos teóricos;
(2) Estimativas com distâncias em rede para os três modelos teóricos; e
(3) Estimativas com distâncias euclidianas na direção principal para os três modelos teóricos.

A fim de comparar o desempenho das estimativas do primeiro caso com as do segundo, calcularam-se 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.

Um segundo método de comparação estabelecido foi a utilização de testes de hipótese não-paramétricos para amostras em pares para comparação de valores típicos (teste da mediana) e distribuições populacionais (testes de Mann-Whitney (M-W) e Kolmogorov-Smirnov (K-S)). Tais testes buscam verificar se há diferença significativa entre dois grupos de valores, confrontando as características de suas respectivas distribuições de probabilidade. Essa análise parte da suposição de que as estimativas emparelhadas possuem a mesma mediana, para o caso do teste da mediana, e igual distribuição de probabilidade, para o caso de M-W e K-S. De maneira simplificada, a significância (valor *p*) correspondente à estatística do teste deve ser maior do que 0,005, considerando-se um nível de confiança de 95%, para que as hipóteses nulas

sejam retidas e o teste acuse a não ocorrência de diferenças significativas entre as estimativas confrontadas.

Nesse âmbito, realizaram-se as seguintes séries de comparação:

1) valores observados *versus* estimativas provenientes de distâncias euclidianas e semivariogramas omnidirecionais;

2) valores observados *versus* estimativas derivadas de distâncias em rede e semivariogramas omnidirecionais;

3) estimativas provenientes de distâncias euclidianas e semivariogramas omnidirecionais *versus* estimativas derivadas de distâncias em rede e semivariogramas omnidirecionais; e

4) estimativas provenientes de distâncias euclidianas e semivariogramas omnidirecionais *versus* estimativas derivadas de distâncias euclidianas e semivariogramas na direção principal.

O primeiro e segundo blocos de comparações, que abrangeu tanto os testes de hipótese quanto as métricas da Tabela 2, permitiu a identificação do modelo teórico que gerava os melhores resultados para as métricas de aderência e valores p dos testes estatísticos. Após encontrado esse melhor semivariograma para as três variáveis de interesse, seguiu-se para a etapa de interpolação espacial por meio da Krigagem Ordinária. Esse estimador geoestatístico é detalhado na seção a seguir.

### 3.2.7 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, o que engloba também as localizações não amostradas. Em síntese, o estimador da Krigagem Ordinária é dado pela Equação 12.

$$Z^*(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) \tag{12}$$

em que $Z^*(x_0)$: valor estimado da variável regionalizada $Z$ na posição geográfica $x_0$;

$\lambda_i$: 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 pouco representativas/deficientes do fenômeno regionalizado fez com que os ponderadores oriundos da krigagem fossem considerados ótimos, isto é, dão origem a valores previstos não enviesados e com variância mínima (MATHERON, 1971). Tal resultado é alcançado a partir da determinação de que a esperança matemática da diferença entre os valores observados e estimados é nula e de que o somatório dos pesos deve ser igual a 1 (YAMAMOTO; 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 números mais precisos.

No que concerne à Krigagem Ordinária (KO), a média local da variável regionalizada ($\mu$), considerada constante pela hipótese de estacionariedade de segunda ordem, também é desconhecida. Consequentemente, além de estimar os ponderadores ótimos, a rotina de cálculos da KO inclui, ainda, a estimativa de $\mu$ (YAMAMOTO; LANDIM, 2015). Outra vantagem dessa técnica é que ela não exige a aquisição de informações adicionais acerca de variáveis explanatórias relacionadas à posição geográfica dos pontos amostrados, apenas os valores da variável de interesse e as coordenadas são necessárias à sua aplicação.

O formalismo estatístico imbuído no estimador da KO, que abrange a estrutura da Equação 12 e as duas restrições mencionadas anteriormente, estabelece o sistema de equações da Krigagem Ordinária em termos da função semivariograma, conforme mostrado na Equação 13 (CRESSIE, 1993; GOOVAERTS, 1997; YAMAMOTO; LANDIM, 2015; MATHERON, 1971).

$$\begin{cases} \sum_{j=1}^{n} \lambda_j \gamma\left(h_{x_i - x_j}\right) + \mu = \gamma\left(h_{x_0 - x_i}\right) \quad para \quad i = 1,n \\ \sum_{j=1}^{n} \lambda_j = 1 \end{cases} \tag{13}$$

Em que $\lambda$ corresponde ao peso ótimo atribuído a cada observação do banco de dados e o termo entre parênteses representa a distância entre as duas localizações $x$. Dessa forma, $\gamma(.)$ expressa a semivariância teórica entre os pares de pontos separados por essa distância. 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.

A superfície contínua de valores interpolados de krigagem foi gerada apenas para as variáveis de Embarques e Desembarques, uma vez que, para Carregamento, há apenas um valor possível para cada segmento de linha, informação que foi atribuída às coordenadas do ponto médio do trecho. No que se refere às estimativas provenientes de distâncias em rede, não se

produziram as superfícies de krigagem para nenhuma das variáveis, pois, nesse caso, as matrizes de distâncias ao longo da linha 856R-10, necessárias à resolução do sistema de equações 13, teriam dimensão impossível de se obter computacionalmente.

Para o cálculo da superfície de krigagem, os parâmetros do semivariograma empírico e teórico foram utilizados como *input* na ferramenta *Geostatistical Analyst*, disponível no *software* ArcGIS 10.1. Tendo em vista os altos recursos computacionais necessários ao cômputo do sistema de equações 13, o método utilizado por essa ferramenta para determinar o tamanho da grade resultou em uma célula de lado aproximadamente igual a 18,7 m, que retorna uma área por quadrícula de 350,8 m². Se considerado um espaçamento médio entre pontos de parada equivalente a 400 m e a dimensão do espaço ocupado por um ponto de parada comum, é possível afirmar que, no presente estudo de caso, tal resolução é aceitável. Tamanhos de grade muito pequenos, apesar de permitirem maior detalhamento do fenômeno, demandam tempo, capacidade e memória computacional bastante elevados, o que pode torna-los impraticáveis.

### 3.2.8 Recursos computacionais

Vale ressaltar que os procedimentos que compõem o método proposto nesse trabalho foram concebidos por meio do apoio de várias ferramentas computacionais. A Tabela 3 relaciona cada etapa do processo ao recurso utilizado para o seu cumprimento, bem como o material bibliográfico auxiliar e/ou a fonte do programa. Os resultados são apresentados no Capítulo 4.

Tabela 3 - Recursos computacionais utilizados no estudo

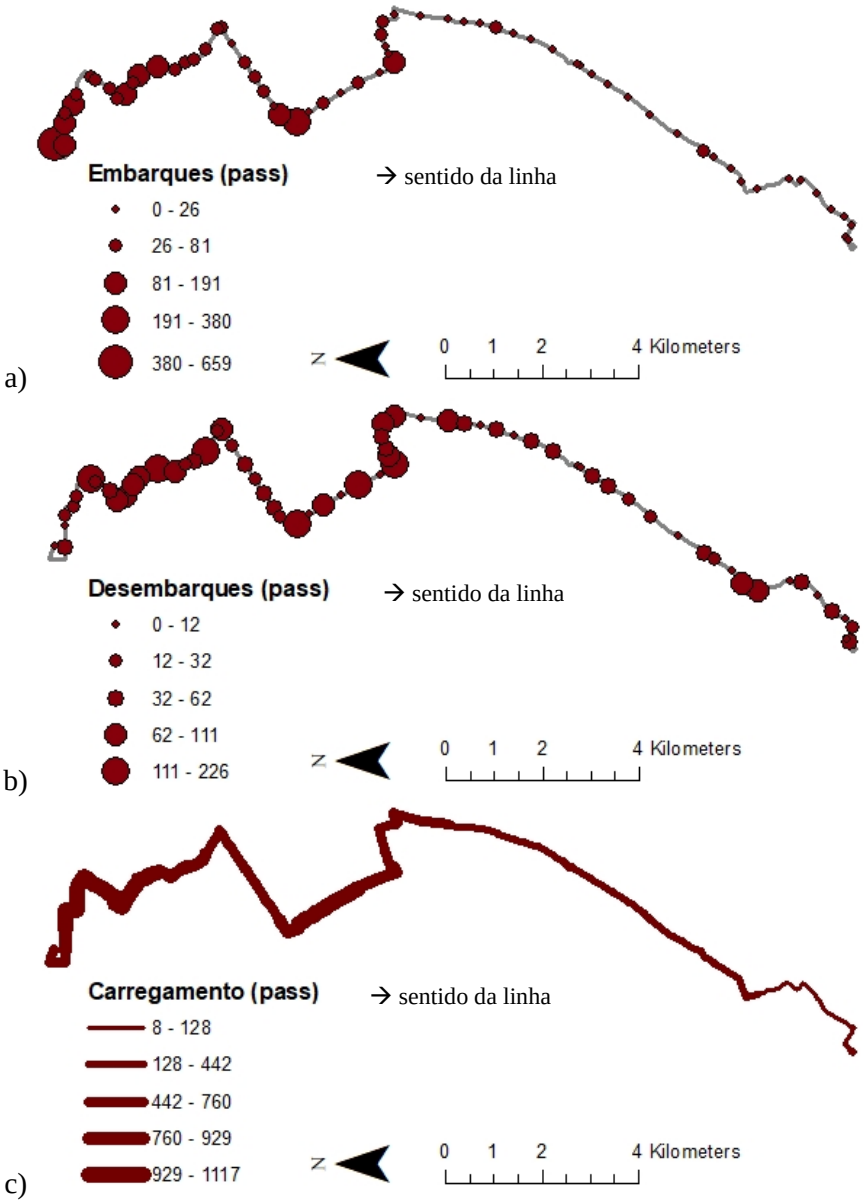
| Etapa | Programa | Material |
|---|---|---|
| Análise exploratória tradicional | IBM SPSS 24.0 | IBM (2016) |
| Transformação de Box-Cox | R/biblioteca EnvStats | R Core Team (2018); Millard (2013) |
| Análise exploratória de dependência espacial | R/biblioteca APE | Paradis, Claude e Strimmer (2004) |
| Análise de autovalores mínimos | R | Ver Hoef (2018) |
| Obtenção das matrizes de distâncias euclidianas | QGIS 2.18.12 | - |
| Obtenção das matrizes de distâncias ao longo da linha de ônibus | GRASS GISS | Bundala, Bergenheim e Metz (2014) |
| Tratamento dos dados e geração da superfície contínua de krigagem | ArcGIS 10.1 | ESRI (2010) |
| Identificação da direção principal e modelagem geoestatística | R/biblioteca geoR | Ribeiro Jr e Diggle (2016); Ver Hoef (2018) |

Fonte: Elaboração própria

# 4 RESULTADOS E DISCUSSÕES

A Figura 5 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.

Figura 5 – Perfis de a) Embarques, b) Desembarques e c) Carregamento na linha 856R-10




Fonte: Elaboração própria

A Figura 5 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 26km.

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.

A Tabela 4 consolida as medidas de estatística descritiva tanto para as variáveis de interesse em sua forma bruta quanto transformadas pelo método de Box-Cox.

Tabela 4 - 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 |

Fonte: Elaboração própria

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. Um outro fator a ser observado é a discrepância do desvio padrão de Embarques quando comparado à sua média. O valor elevado encontrado para essa medida de dispersão, antes da transformação, destaca a ocorrência de eventuais *outliers* no banco de dados e sua grande amplitude de variação (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 gerados de viagens terão um número de Embarques muito maior que os vizinhos, o que contribui para a elevação do desvio padrão da variável.

Além disso, ressalta-se também que, conforme mostrado na Tabela 5, o coeficiente de correlação do gráfico de probabilidade (PPCC, *Probability Plot Correlation Coefficient*) das três variáveis aumentou após a aplicação da potência ótima proveniente da transformação de Box-Cox.

Tabela 5 – Potências ótimas da transformação de Box-Cox

| Variável | Potência ótima | PPCC antes da transformação | PPCC depois da transformação |
|---|---|---|---|
| Embarques | 0,1070194 | 0,69680 | 0,9883505 |
| Desembarques | 0,2501787 | 0,90214 | 0,9925090 |
| Carregamento | 1,1517130 | 0,96608 | 0,9665636 |

Fonte: Elaboração própria

Apesar de as três variáveis já demonstrarem, antes da transformação, um valor razoável para o PPCC (que avalia a proximidade entre os dados e a distribuição normal), destaca-se que, antes da aplicação da potência ótima de Box-Cox, nenhuma delas 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 realizada a Análise Espacial Exploratória a partir do cálculo do indicador global de associação espacial índice de Moran. A Tabela 6 sintetiza os resultados para o índice de Moran, calculados para as variáveis de interesse transformadas e sua respectiva significância.

Tabela 6 - Autocorrelação espacial das variáveis de interesse

| Variável | Tipo de distância | Moran's I | Valor Esperado | Padrão | Valor p | Desvio Padrão |
|---|---|---|---|---|---|---|
| Embarques | euclidiana | 0,4871466 | -0,01785714 | Agrupado | 0 | 0,04537344 |
| | em rede | 0,4439056 | -0,01785714 | Agrupado | 0 | 0,04565837 |
| Desembarques | euclidiana | 0,1731550 | -0,01785714 | Agrupado | 2,82E-05 | 0,04561527 |
| | em rede | 0,1904801 | -0,01785714 | Agrupado | 5,66E-06 | 0,04590162 |
| Carregamento | euclidiana | 0,6344457 | -0,01818182 | Agrupado | 0 | 0,03796773 |
| | em rede | 0,6708997 | -0,01818182 | Agrupado | 0 | 0,04245522 |

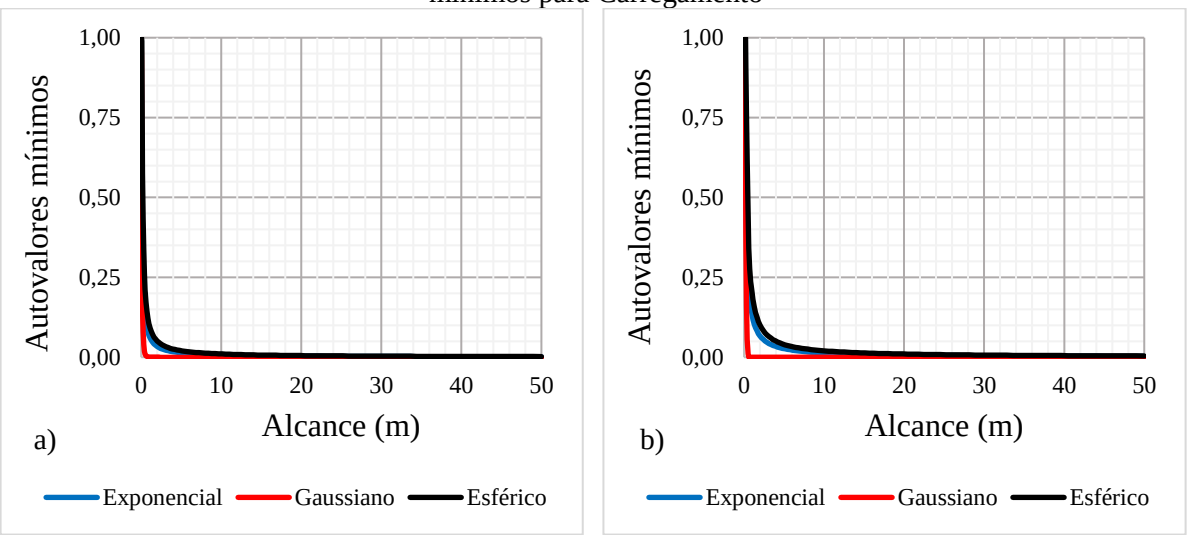
Fonte: Elaboração própria

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. A crescente intensidade no valor do índice a partir de Desembarques, passando por Embarques e encerrando em Carregamento reflete também a continuidade espacial presente nessas funções e pode já dar um indício do desempenho do estimador geoestatístico a ser 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.

Nota-se, também, um aumento no valor da autocorrelação quando os pesos utilizados para o cálculo do índice são referentes ao inverso da distância entre pontos ao longo da linha de ônibus (distâncias em rede). Apenas para a variável de Embarques houve uma queda no valor do indicador, embora a significância do índice tenha se mantido elevada. Tais resultados podem apontar que, apesar de haver certa diferença entre a autocorrelação euclidiana com a de distâncias em rede, tanto para mais quanto para menos, não se tratam de diferenças significativas devido, provavelmente, ao traçado regular da linha, que aproxima visivelmente as distâncias em rede das euclidianas.

A Figura 6 expõe a avaliação dos autovalores mínimos encontrados para o arranjo espacial dos pontos de parada (gráfico à esquerda) e pontos médios dos segmentos de linha (gráfico à direita), ambos associados às distâncias ao longo da linha 856R-10. Percebe-se que, ao aumentar o parâmetro alcance, os autovalores se aproximam assintoticamente de zero, motivo pelo qual os gráficos se estendem somente até uma distância de 50 m.

Figura 6 – a) Autovalores mínimos para Embarques e Desembarques; e b) Autovalores mínimos para Carregamento



Fonte: Elaboração própria

A figura acima mostra que, para nenhum dos modelos teóricos aplicados, houve risco de obtenção de erros padrões e variâncias negativas nas estimativas de krigagem. Mesmo que a faixa de valores para o eixo das abscissas não abranja a provável extensão do alcance para as variáveis de interesse, o comportamento assintótico da curva evidencia a inexistência de autovalores negativos.

O motivo para a não ocorrência de autovalores negativos se deve, provavelmente, ao arranjo relativamente simples da linha 856R-10, que, conforme mencionado anteriormente, torna equivalentes as distâncias em rede e em linha reta. Essa é uma das vantagens de se modelar, separadamente, os percursos de ida e volta, já que o risco de aparecimento de autovalores negativos seria maior em um itinerário fechado. Exemplos de traçados que manifestam essa característica, como, por exemplo, o circular e aqueles em que há sobreposição, confluência e/ou interseção de linhas, foram disponibilizados por Ver Hoef (2018), sendo que todos eles apresentavam uma complexidade não verificada na linha 856R-10, em seu sentido de ida.

## 4.1. Comparação entre estimativas provenientes de distâncias euclidianas e em rede

Esta seção descreve os resultados provenientes de duas comparações: (1) Valores reais *versus* valores estimados a partir de semivariogramas omnidirecionais de distâncias euclidianas; e (2) Valores reais *versus* valores estimados a partir de semivariogramas omnidirecionais de distâncias em rede. Foram testados três tipos de curvas de semivariogramas em ambos os casos. Como se tratam de três variáveis, nove resultados são apresentados na Tabela 7.

A Tabela 7 contém as métricas de erros para os três tipos de semivariogramas considerados para as variáveis de interesse transformadas pelo método de Box-Cox. Cabe lembrar que tais resultados se referem à comparação entre os valores reais e aqueles obtidos por meio do teste do ponto fictício.

Tabela 7 - Desempenho de três modelos de semivariograma à estimativa das variáveis

| Variável | Modelo teórico | Tipo de distância | SE | ME | MAE | RMSE | R |
|---|---|---|---|---|---|---|---|
| Embarques | Exponencial | em rede | 493.126,50 | -18,16 | 34,57 | 93,01 | 0,432** |
| | | euclidiana | 450.573,94 | -15,65 | 35,01 | 88,91 | 0,497** |
| | Gaussiano | em rede | 485.466,53 | -18,82 | 34,13 | 92,29 | 0,469** |
| | | euclidiana | 476.718,32 | -17,97 | 33,91 | 91,45 | 0,477** |
| | Esférico | em rede | 494.996,00 | -18,85 | 34,41 | 93,19 | 0,439** |
| | | euclidiana | 462.663,01 | -17,10 | 34,01 | 90,09 | 0,491** |
| Desembarques | Exponencial | em rede | 147.748,64 | -11,41 | 32,15 | 50,91 | 0,389** |
| | | euclidiana | 152.778,03 | -12,80 | 32,88 | 51,77 | 0,360** |
| | Gaussiano | em rede | 140.377,34 | -10,98 | 31,15 | 49,63 | 0,435** |
| | | euclidiana | 149.215,23 | -12,54 | 32,63 | 51,16 | 0,387** |
| | Esférico | em rede | 141.118,15 | -11,13 | 31,19 | 49,76 | 0,432** |
| | | euclidiana | 149.853,29 | -12,58 | 32,62 | 51,27 | 0,382** |
| Carregamento | Exponencial | em rede | 80.251,81 | 0,61 | 24,65 | 37,86 | 0,993** |
| | | euclidiana | 102.417,34 | 0,00 | 27,11 | 42,77 | 0,992** |
| | Gaussiano | em rede | 196.842,35 | 2,10 | 42,55 | 59,29 | 0,984** |
| | | euclidiana | 297.844,83 | 2,81 | 51,30 | 72,93 | 0,975** |
| | Esférico | em rede | 76.593,75 | 0,46 | 24,77 | 36,98 | 0,994** |
| | | euclidiana | 100.987,47 | -0,16 | 27,06 | 42,47 | 0,992** |

Fonte: Elaboração própria
Nota: ** A correlação é significativa ao nível de confiança de 99% (1 extremidade). SE, ME, MAE, RMSE e R são, respectivamente, Erro Quadrado, Erro Médio, Erro Médio Absoluto, Raiz Quadrada do Erro Quadrado Médio e Coeficiente de Correlação Linear de *Pearson*.

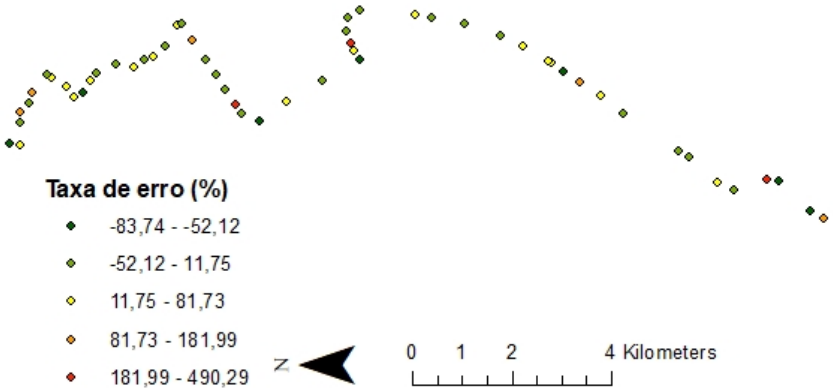
É importante ressaltar que, apesar de a modelagem dos semivariogramas ótimos ter sido realizada a partir da validação cruzada com os valores das variáveis transformadas, os erros catalogados na Tabela 7 foram obtidos por meio dos valores reais e estimados já resultantes da transformação inversa. Conforme 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 significativos, fortalecendo a conclusão de que, certamente, há uma razoável semelhança entre os valores reais observados 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 alguns semivariogramas 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

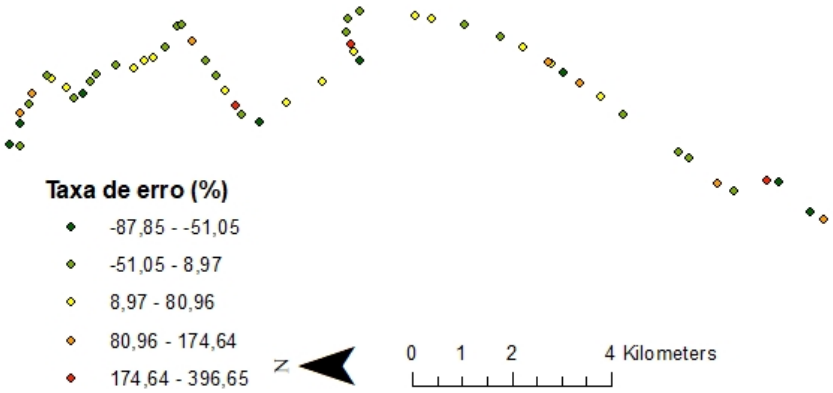
modelos de que tiveram predominância de menores desvios e maiores coeficientes de correlação linear foram o exponencial, para o número de Embarques, o gaussiano, para o volume de Desembarques e o esférico, para Carregamento, em se tratando de distâncias euclidianas. No que se refere ao semivariogramas com distâncias em rede, se destacaram o modelo gaussiano, para Embarques e Desembarques, e novamente o esférico, para Carregamento. A fim de prover subsídios a uma análise pontual dos erros, os mapas disponibilizados na Figura 7 expõem as taxas de desvio que as estimativas de Embarques e Desembarques geradas pelos semivariogramas supracitados apresentaram com relação ao valor observado da respectiva variável em cada ponto de parada.

Haja vista que as porcentagens de erro dependem do número real de Embarques e Desembarques, não foi possível calculá-las para os pontos 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.

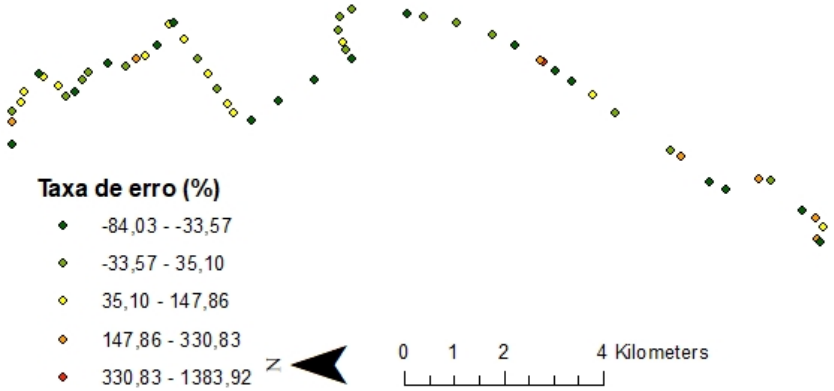
Figura 7 – Porcentagem de desvio nas estimativas de krigagem para Embarques e Desembarques



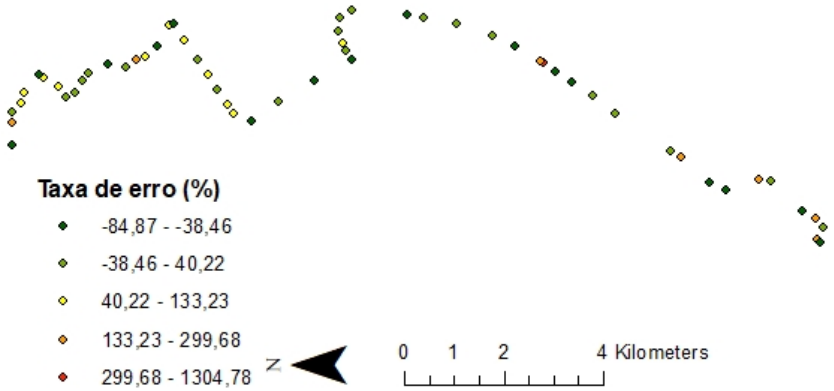

a) Embarques com distâncias euclidianas e semivariograma exponencial




b) Embarques com distâncias em rede e semivariograma gaussiano




c) Desembarques com distâncias euclidianas e semivariograma gaussiano




d) Desembarques com distâncias em rede e semivariograma gaussiano

Fonte: Elaboração própria

De uma forma geral, quanto mais escura (e/ou forte) for a cor do ponto, tanto para verde quanto vermelho, maior é a taxa de desvio do valor estimado ao observado. De acordo com essa especificação, percebe-se que, para as quatro estimativas, houve mais desvios negativos do que positivos, ou seja, a modelagem geoestatística previu, na maioria, um número de Embarques e de Desembarques inferior aos valores coletados na pesquisa sobe/desce. Tal constatação corrobora a ocorrência de erros médios negativos comentada anteriormente.

Cabe destacar que, enquanto as porcentagens máximas de subestimação se limitam à ordem de 80%, no caso de superestimação, essa taxa aumenta para aproximadamente 400% e 1.300% em Embarques e Desembarques, respectivamente, o que ratifica o melhor desempenho da primeira variável em comparação com a segunda. Entretanto, tendo em vista a quantidade ínfima de pontos que se enquadraram na última faixa de erros (de cor vermelha), é possível afirmar que tanto Embarques quanto Desembarques demonstram resultados satisfatórios também quando analisados os erros por cada ponto de parada.

Outro 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 quatro casos apresentados na Figura 7: 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.

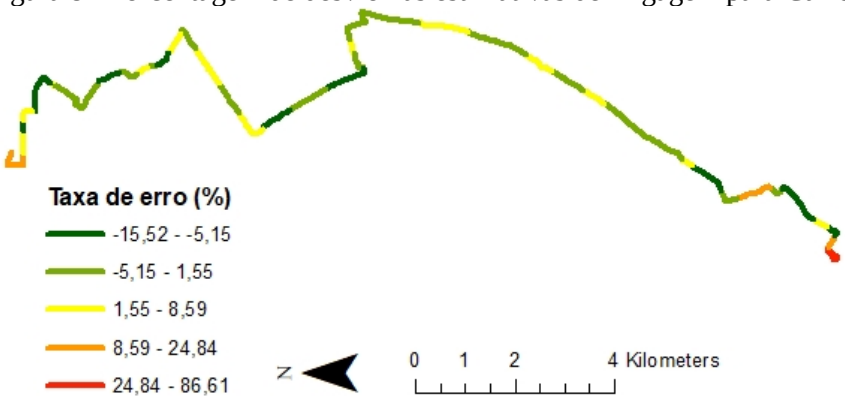
Conhecendo o perfil das três variáveis de interesse, é possível afirmar que a ocorrência de erros extremos, no caso de Embarques e Desembarques, é devida ao fato de tais funções não serem tão suaves, espacialmente, quanto Carregamento. A sobreposição dos mapas da Figura 7 com os da Figura 5 revela que as maiores taxas de desvio, que contém 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 é 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.

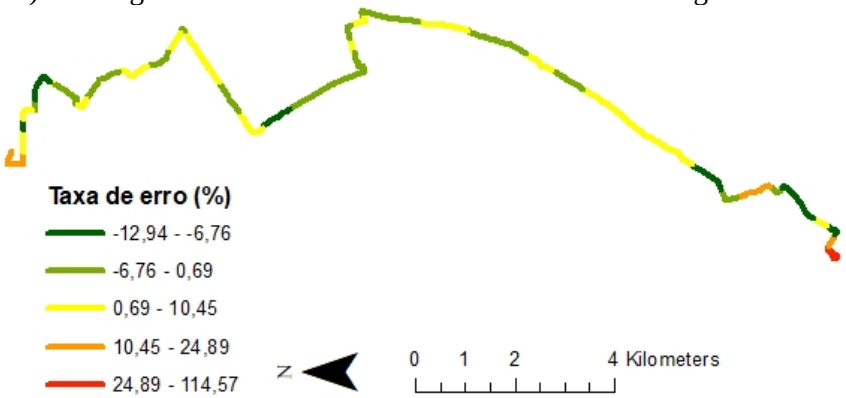
Os mapas da Figura 7 corroboram a semelhança entre as métricas de aderência encontradas para as estimativas de distâncias em rede e euclidianas. Percebe-se que os limites das categorias permanecem quase inalterados e os pontos em cada classe de erro praticamente se repetem nos dois tipos de distância, inclusive no que se refere aos grupos de erros extremos.

A Figura 8 mostra a variação de erros para a variável de Carregamento. Apesar dos mapas a) e b) apresentarem as porcentagens de desvio por trecho, tal representação foi escolhida apenas para permitir uma melhor visualização e facilitar a interpretação dos resultados. Porém, conforme descrito na etapa de tratamento dos dados (Seção 3.2.1), o Carregamento em cada segmento da linha foi atribuído ao ponto médio de cada ramo para a realização da modelagem geoestatística. Esses pontos estão discriminados no mapa c).

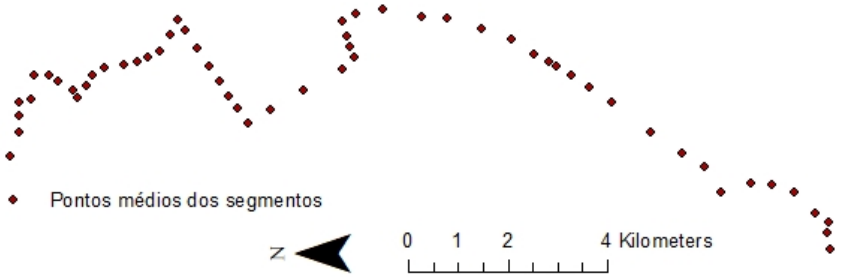
Figura 8 - Porcentagem de desvio nas estimativas de krigagem para Carregamento




a) Carregamento com distâncias euclidianas e semivariograma esférico




b) Carregamento com distâncias em rede e semivariograma esférico




c) Pontos médios dos trechos que receberam os valores de Carregamento

Fonte: Elaboração própria

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 exibe um excelente desempenho: os erros se mantêm entre -5% e 10%, 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.

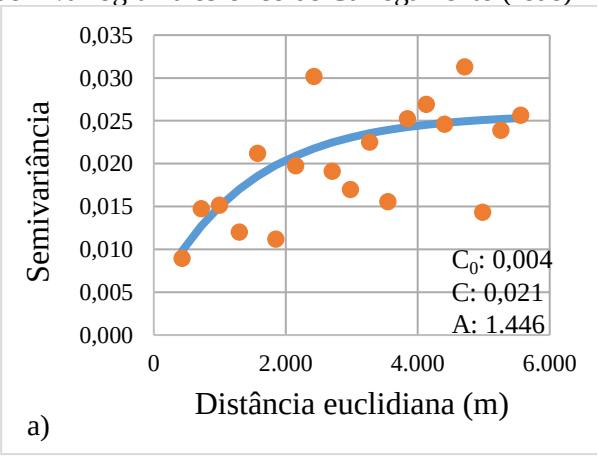
As taxas de desvio de Embarques, Desembarques e Carregamento apresentaram uma assimetria positiva também verificada em resultados de trabalhos anteriores (CHI; ZHENG,

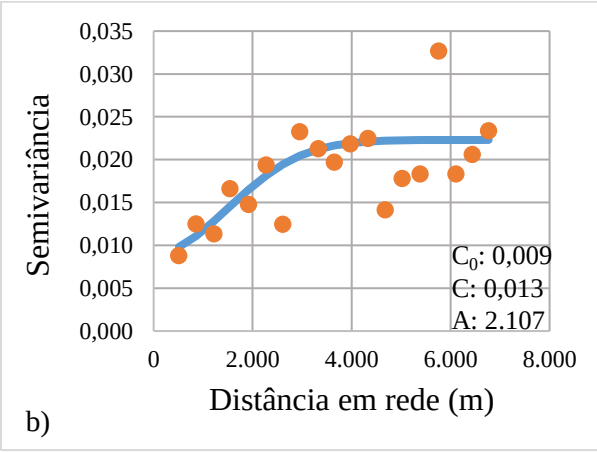
2013; WANG; KOCKELMAN, 2009). A mediana dos erros em porcentagem foi de -1,42%, -1,72% e 0,37%, respectivamente, para Embarques, Desembarques e Carregamento nos casos da Figura 7 e Figura 8, o que confirma, mais uma vez, a tendência de subestimação do estimador de krigagem. Em se tratando dos desvios absolutos, as três variáveis demonstraram uma mediana de aproximadamente 44%, 48% e 3%, e média de 71%, 92% e 7%.

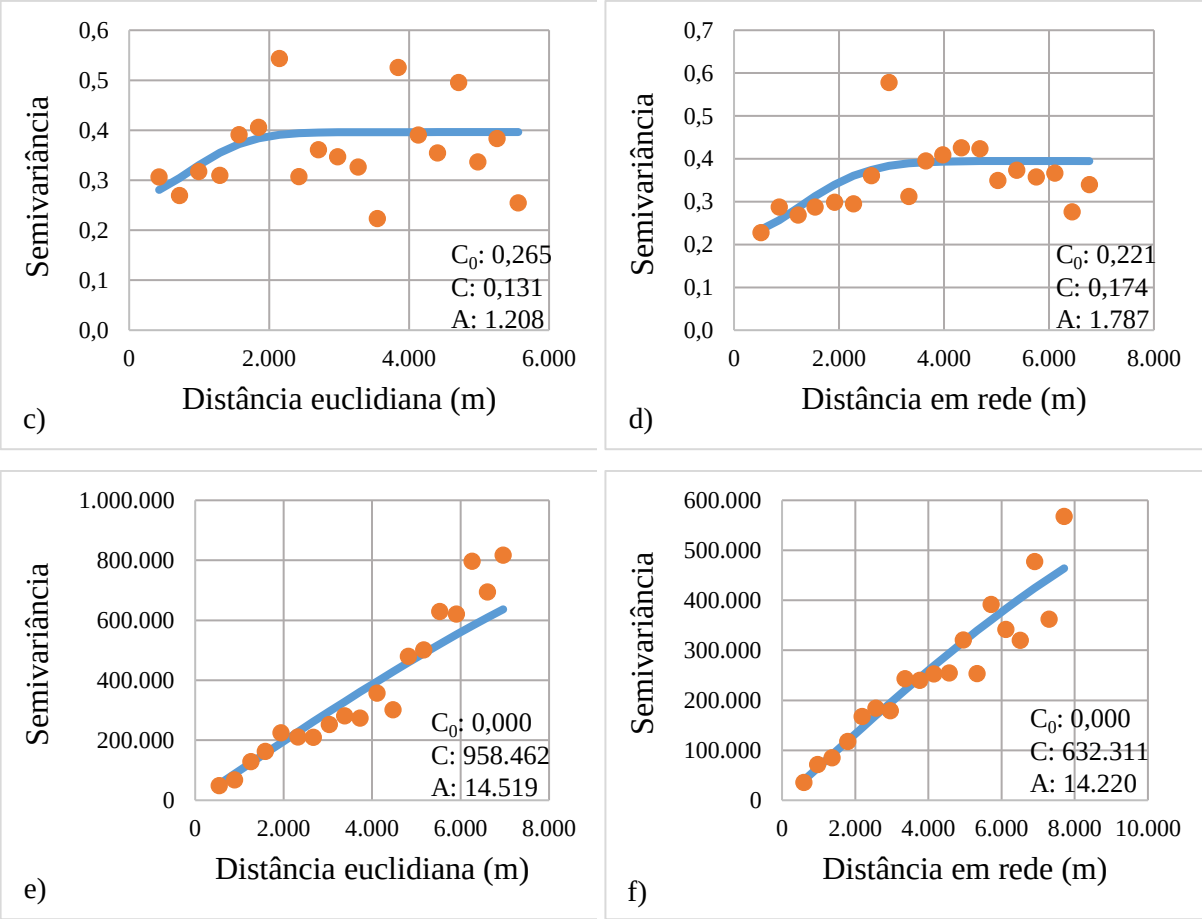
Apesar de os resultados obtidos para Embarques e Desembarques parecerem desanimadores, em comparação com estudos que utilizaram a KO com validação cruzada (CHI; ZHENG, 2013; SHAMO et al., 2015), tais valores podem ser reconhecidos como satisfatórios. A média de erros encontrada por Chi e Zheng (2013) ficou em torno de 59%, porém, o menor valor de RMSE calculado pelos autores foi de 124. Cabe lembrar que o maior RMSE obtido no presente trabalho é da ordem de 93, conforme mostrado na Tabela 7. 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 89, 50 e 37.

No intuito de visualizar o ajuste das curvas teóricas às semivariâncias médias entre pares, os semivariogramas de melhor desempenho estão mostrados na Figura 9 juntamente com os parâmetros exatos dos modelos calibrados. Os demais modelos teóricos estão disponíveis na Figura 16 do Apêndice.

Figura 9 - a) Semivariograma exponencial de Embarques (euclidianas); b) Semivariograma gaussiano de Embarques (rede); c) Semivariograma gaussiano de Desembarques (euclidianas); d) Semivariograma gaussiano de Desembarques (rede); e) Semivariograma esférico de Carregamento (euclidianas); e f) Semivariograma esférico de Carregamento (rede)






Fonte: Elaboração própria

No que tange aos parâmetros mostrados na Figura 9, destaca-se os valores de efeito pepita encontrado para as três variáveis de interesse. Em termos relativos, no que concerne aos semivariogramas em linha reta, Embarques demonstrou uma descontinuidade na origem de, aproximadamente, 16,67% do patamar (semivariância máxima, $C_0$ + C), o que certamente contribuiu para um melhor desempenho dessa variável em comparação com a de Desembarques, cujo efeito pepita foi de 66,91% do patamar. Em se tratando dos semivariogramas em rede, a aleatoriedade na vizinhança imediata aumentou para 40,60%, no caso de Embarques, e diminuiu para 55,99% em Desembarques.

Embora, à primeira vista, essa alta 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 de Desembarques variável foi de 29,91%, correspondente ao modelo exponencial com distâncias em rede e bem mais próximo do limite de 30% do que aqueles relativos aos modelos gaussianos. Porém, tendo em vista que as métricas de aderência para o segundo modelo foram melhores que as 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 variáveis explicativas 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.

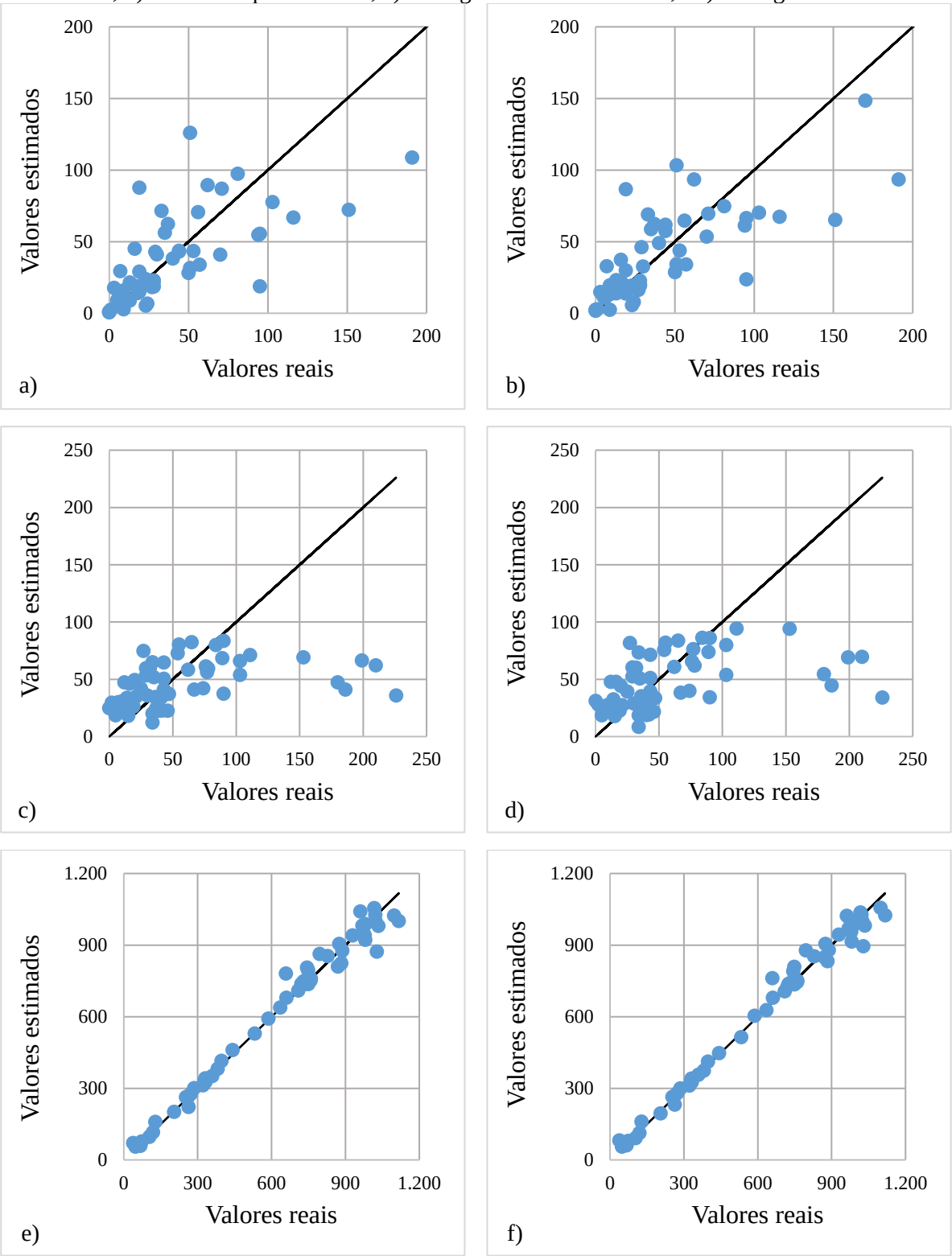
Quando comparados os mesmos modelos teóricos ajustados aos semivariogramas com distâncias em linha reta e ao longo da linha de ônibus para as variáveis de Embarques e Desembarques, é possível afirmar que, de uma forma geral, a utilização de distâncias em rede contribuiu para a suavização da curva calibrada aos pontos dos semivariogramas experimentais. Apenas o modelo exponencial, em se tratando de Embarques, sofreu um aumento na porcentagem de efeito pepita relativa ao patamar, sendo que para todas as outras curvas houve redução na descontinuidade na origem, algumas delas bastante expressivas, como no caso do modelo exponencial de Desembarques, que diminuiu de 51,51% com distâncias euclidianas para 29,91% com distâncias em rede.

É importante destacar que Desembarques apresentou um desempenho inferior ao de Embarques apenas 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 a variável de Desembarques foram menores que para Embarques, o que mostra uma aderência considerável do volume estimado de passageiros que descem ao real coletado pela pesquisa sobe/desce.

Por fim, destaca-se o excelente desempenho alcançado para a variável Carregamento. Possuindo um efeito pepita praticamente desprezível em todas as combinações de modelos, exceto o gaussiano, e tipos de distância, suas métricas de erros foram 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. Apenas o modelo gaussiano gerou aleatoriedade na origem diferente do zero virtual, o que, mesmo assim, não impediu que os resultados provenientes dessa curva fossem ainda superiores aos de Embarques e Desembarques. A Figura 10 ilustra os comentários

realizados anteriormente ao mostrar a aderência das estimativas provenientes da Krigagem Ordinária à reta de 45°, na qual os valores estimados são exatamente iguais aos observados.

Figura 10 – Desempenho de a) Embarques em linha reta; b) Embarques em rede; c) Desembarques em linha reta; d) Desembarques em rede; e) Carregamento em linha reta; e f) Carregamento em rede




Fonte: Elaboração própria

A comparação dos gráficos a) com c) e b) com d) permite inferir o motivo pelo qual as métricas de erros resultaram maiores para Desembarques que para Embarques, porém menor R para a primeira em confronto com a segunda. Percebe-se que, para a faixa de valores de 0 a aproximadamente 125, as estimativas de passageiros que descem se mantêm mais aderidas à reta de 45° se comparadas ao caso dos passageiros que sobem, que demonstra uma notável dispersão nessa mesma janela de pontos. Em contraposição, a partir de 150 passageiros que desembarcam, os pontos de valores reais por valores estimados se afastam substancialmente da bissetriz, o que provavelmente contribuiu para a obtenção de um coeficiente de correlação, para essa variável, inferior ao de Embarques.

Nota-se que, apesar de Carregamento ter resultado nas melhores métricas de aderência na validação cruzada, o alcance obtido para essa variável foi consideravelmente grande, aproximadamente igual à distância máxima média entre pontos do banco de dados. Esse comportamento normalmente é visualizado em variáveis regionalizadas que possuem tendências espaciais não modeladas no processo de krigagem. Tal fenômeno mascara o verdadeiro comportamento espacial da variável, resultando em parâmetros do semivariograma teórico equivocados. A exemplo de Desembarques, a eliminação desse viés se dá por meio da adição de variáveis explicativas/independentes ao interpolador, artifício ignorado no presente trabalho por razões já expostas.

A Tabela 8 mostra a significância dos testes de hipóteses aplicados à comparação entre as estimativas obtidas pelo teste do ponto fictício considerando-se os semivariogramas com distâncias em linha reta e aquelas geradas pelas semivariâncias ao longo da linha de ônibus. Nesse âmbito, as prováveis diferenças significativas poderão ser atribuídas a um melhor ou pior desempenho da respectiva combinação variável x modelo teórico x teste para o caso de distâncias em rede em confronto com as euclidianas.

Tabela 8 – Testes de hipóteses não-paramétricos aplicados à comparação de resultados entre Krigagem Ordinária com distâncias euclidianas e com distâncias em rede

| Variável | Modelo teórico | Teste | Valor p | Conclusão |
|---|---|---|---|---|
| Embarques (Em rede x Euclidiana) | Exponencial | Mediana | 1,000 | Não há diferenças significativas |
| | | Mann-Whitney | 0,966 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 1,000 | Não há diferenças significativas |
| | Gaussiano | Mediana | 1,000 | Não há diferenças significativas |
| | | Mann-Whitney | 0,975 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,999 | Não há diferenças significativas |
| | Esférico | Mediana | 1,000 | Não há diferenças significativas |
| | | Mann-Whitney | 0,939 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 1,000 | Não há diferenças significativas |
| Desembarques (Em rede x Euclidiana) | Exponencial | Mediana | 1,000 | Não há diferenças significativas |
| | | Mann-Whitney | 0,908 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,476 | Não há diferenças significativas |
| | Gaussiano | Mediana | 1,000 | Não há diferenças significativas |
| | | Mann-Whitney | 0,899 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,628 | Não há diferenças significativas |
| | Esférico | Mediana | 1,000 | Não há diferenças significativas |
| | | Mann-Whitney | 0,899 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,476 | Não há diferenças significativas |
| Carregamento (Em rede x Euclidiana) | Exponencial | Mediana | 0,850 | Não há diferenças significativas |
| | | Mann-Whitney | 1,000 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 1,000 | Não há diferenças significativas |
| | Gaussiano | Mediana | 0,850 | Não há diferenças significativas |
| | | Mann-Whitney | 0,921 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,979 | Não há diferenças significativas |
| | Esférico | Mediana | 0,850 | Não há diferenças significativas |
| | | Mann-Whitney | 0,995 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 1,000 | Não há diferenças significativas |

Fonte: Elaboração própria

Observa-se, a partir da Tabela 8, que não há diferenças significativas, tanto em termos de valores típicos (teste da mediana), quanto em termos de distribuição populacional (Mann-Whitney e Kolmogorov-Smirnov) entre valores estimados a partir de KO (variogramas ominidirecionais) para distâncias em rede e distâncias euclidianas. Isso se deve, provavelmente, ao fato de que a linha 856R-10 possui um traçado relativamente simples no que diz respeito à sua retidão, o que aproxima, nesse caso, as distâncias em rede das euclidianas. Consequentemente, os pares encontrados para o cálculo das semivariâncias médias de acordo com os *lags* e tolerâncias definidos acabam se repetindo em ambos os tipos de distância. Dessa forma, os parâmetros calibrados dos semivariogramas teóricos resultam semelhantes nos dois casos.

Cabe resgatar, nesse estágio da discussão, os resultados da comparação, também por meio de testes de hipóteses, do volume observado de Embarques, Desembarques e Carregamento com seus respectivos valores estimados, sendo que tal análise fora dividida em dois blocos, conforme definido na seção 3.2.6: 1) quantidades observadas e estimadas pela KO com distâncias euclidianas; e 2) quantidades observadas e estimadas pela KO com distâncias em rede. A significância dos testes realizados, cujos valores p estão dispostos na Tabela 10 e Tabela 11 do Apêndice e não serão exibidas no presente tópico por brevidade, corroboram as conclusões apontadas na Tabela 8, ou seja, para todas as combinações de testes, modelos teóricos e variáveis, os números contidos nos dois grupos de comparações não apresentaram diferenças significativas. Tais resultados indicam a capacidade de ajustes de todos os modelos, pois apontam similaridade entre valores estimados e observados para todos os casos propostos.

Essa verificação traz à tona duas inferências principais, a saber: tanto a utilização de distâncias euclidianas quanto em rede conseguiu alcançar estimativas notadamente próximas dos valores observados de Embarques e Desembarques por ponto de parada e Carregamento por segmento da linha 856R-10. A segunda conclusão é a de que, apesar das diferenças entre os dois tipos de estimativas no que tange às métricas de aderência, não há indícios, no estudo de caso realizado, de que as distâncias em rede devam ser priorizadas em detrimento das euclidianas por apresentarem melhores resultados.

Embora Desembarques em rede e Carregamento em rede tenham demonstrado um desempenho superior aos de distâncias em linha reta, enquanto que para Embarques aconteceu o contrário, os testes de hipóteses evidenciam/esclarecem que tais variações nas estimativas provenientes de um tipo de distância para as geradas pelo outro não são significativas. Dessa forma, não se pode provar, apenas com o presente estudo de caso, que há ganhos substanciais quando são utilizadas distâncias em rede em detrimento das euclidianas.

O Quadro 1 resume os principais resultados comparativos relativos à estimativa geoestatística a partir de distâncias em rede e distâncias euclidianas. Vale ressaltar que os valores de referência, para comparações, são os valores observados.

Quadro 1 – Resumo da comparação entre estimativas provenientes de distâncias em rede e euclidianas

| Variáveis | Melhor modelo variográfico | Métricas de aderência | % do efeito pepita | Testes não-paramétricos |
|---|---|---|---|---|
| Embarques (rede) | Gaussiano | Pequenos valores de erros e boa correlação | 40,60 | Apontaram similaridades de mediana e distribuição entre abordagens em rede e euclidianas |
| Embarques (euclidianas) | Exponencial | Erros inferiores ao de Embarques em rede e correlação um pouco maior que a desta abordagem | 16,67 | |
| Desembarques (rede) | Gaussiano | Erros inferiores aos de Desembarques em linha reta e correlação um pouco maior que a desta abordagem | 55,99 | |
| Desembarques (euclidianas) | Gaussiano | Erros inferiores aos das duas abordagens de Embarques e uma boa correlação | 66,91 | |
| Carregamento (rede) | Esférico | Erros inferiores aos das duas abordagens de Embarques e Desembarques e de Carregamento em linha reta e alta correlação | 0,00 | |
| Carregamento (euclidianas) | Esférico | Erros inferiores aos das duas abordagens de Embarques e Desembarques e alta correlação | 0,00 | |

Fonte: Elaboração própria

## 4.2. Comparação entre estimativas provenientes de distâncias euclidianas com semivariogramas omnidirecionais e na direção principal

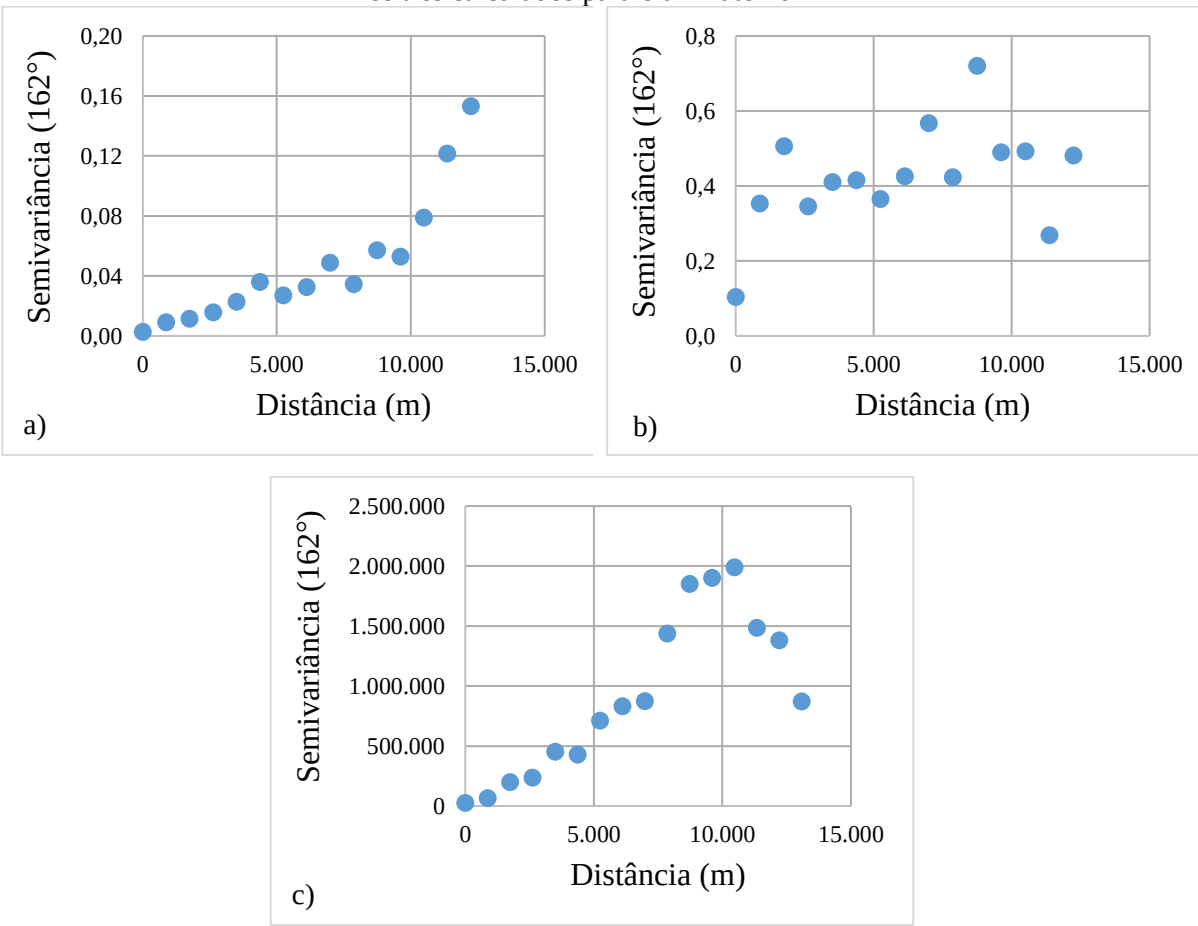
Esta seção descreve os resultados provenientes das seguintes comparações: estimativas provenientes de distâncias euclidianas e semivariogramas omnidirecionais *versus* estimativas derivadas de distâncias euclidianas e semivariogramas na direção principal.

No que tange à identificação da direção principal do fenômeno regionalizado, três direções, das dez que foram testadas, se destacaram quanto aos critérios estabelecidos na Seção

3.2.4 na seguinte ordem: 180°, 18° e 162°. Adotou-se, para efeito de comparação com os resultados do semivariograma omnidirecional, a orientação de 162°, uma vez que, pelo menos um dos três modelos teóricos ajustados aos semivariogramas dos dois primeiros azimutes deu origem a matrizes de semivariância singulares, isto é, não invertíveis, na etapa de validação cruzada. Dessa forma, não foi possível gerar as estimativas do teste do ponto fictício para tais direções.

A Figura 11 mostra os semivariogramas experimentais de Embarques, Desembarques e Carregamento na direção de 162°, sendo que os gráficos obtidos para os outros nove azimutes estão disponíveis na Figura 13, Figura 14 e Figura 15 do Apêndice. É importante relembrar que, para a geração da superfície contínua de krigagem, são necessários tanto o semivariograma na direção principal quanto no azimute perpendicular a essa direção. Uma vez que a linha 856R-10 possui uma orientação bastante aparente, é possível afirmar que, nesse caso, não existiria pontos no semivariograma da direção ortogonal e, portanto, o cálculo da superfície de krigagem seria inconcebível.

Figura 11 –Semivariograma empírico de a) Embarques, b) Desembarques e c) Carregamento, os três calculados para o azimute 162°




Fonte: Elaboração própria

De acordo com a Figura 11, apenas Carregamento não manteve um perfil ascendente até alcançar o patamar, ordenada equivalente à semivariância máxima do banco de dados. Ao invés disso, a partir de uma distância média entre pares de 10.000m, a semivariância de Carregamento começa a diminuir. Porém, alerta-se que o alcance da variável se situou aproximadamente nesse limite e, tendo em vista que as estimativas de krigagem dependem muito mais das observações localizadas na vizinhança próxima do ponto a ser interpolado, tal comportamento não afeta os resultados.

Além disso, cabe mencionar que pontos distanciados de 10.000m ou mais recebem pesos nulos quando da resolução do sistema de equações 13, também baseado no conceito do parâmetro de alcance, descrito na Seção 3.2.5. A Tabela 9 apresenta os valores p dos testes aplicados à comparação das estimativas euclidianas omnidirecionais com aquelas obtidas a partir do azimute 162°. O objetivo é inferir se haveria diferenças significativas entre os valores estimados da Krigagem Ordinária com distâncias em rede omnidirecionais e na direção principal, caso fosse possível calculá-los para esse segundo caso.

Tabela 9 - Testes de hipóteses não-paramétricos aplicados à comparação de resultados entre Krigagem Ordinária omnidirecional e na direção de 162°, ambas com distâncias euclidianas

| Variável | Modelo teórico | Teste | Valor p | Conclusão |
|---|---|---|---|---|
| Embarques | Exponencial | Mediana | 0,851 | Não há diferenças significativas |
| | | Mann-Whitney | 1,000 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 1,000 | Não há diferenças significativas |
| | Gaussiano | Mediana | 1,000 | Não há diferenças significativas |
| | | Mann-Whitney | 0,984 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 1,000 | Não há diferenças significativas |
| | Esférico | Mediana | 0,851 | Não há diferenças significativas |
| | | Mann-Whitney | 1,000 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 1,000 | Não há diferenças significativas |
| Desembarques | Exponencial | Mediana | 0,454 | Não há diferenças significativas |
| | | Mann-Whitney | 0,863 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,628 | Não há diferenças significativas |
| | Gaussiano | Mediana | 0,708 | Não há diferenças significativas |
| | | Mann-Whitney | 0,980 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,239 | Não há diferenças significativas |
| | Esférico | Mediana | 1,000 | Não há diferenças significativas |
| | | Mann-Whitney | 0,962 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,910 | Não há diferenças significativas |
| Carregamento | Exponencial | Mediana | 0,850 | Não há diferenças significativas |
| | | Mann-Whitney | 0,958 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 1,000 | Não há diferenças significativas |
| | Gaussiano | Mediana | 0,850 | Não há diferenças significativas |
| | | Mann-Whitney | 0,981 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,979 | Não há diferenças significativas |
| | Esférico | Mediana | 0,850 | Não há diferenças significativas |
| | | Mann-Whitney | 0,995 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 1,000 | Não há diferenças significativas |

Fonte: Elaboração própria

A Tabela 9 revela que, para todas as combinações de modelos, variáveis e testes, o valor *p* aponta a inexistência de indícios que permitam a rejeição da hipótese nula, isto é, não há evidência de diferenças significativas entre a mediana e distribuição do conjunto de estimativas euclidianas omnidirecionais em comparação com as do azimute 162°. Essa conclusão permite argumentar que, se fosse possível calcular semivariogramas empíricos com distâncias em rede na direção principal, as estimativas provenientes destes não seriam significativamente diferentes, para mais ou para menos, daquelas geradas pelos semivariogramas omnidirecionais também em rede.

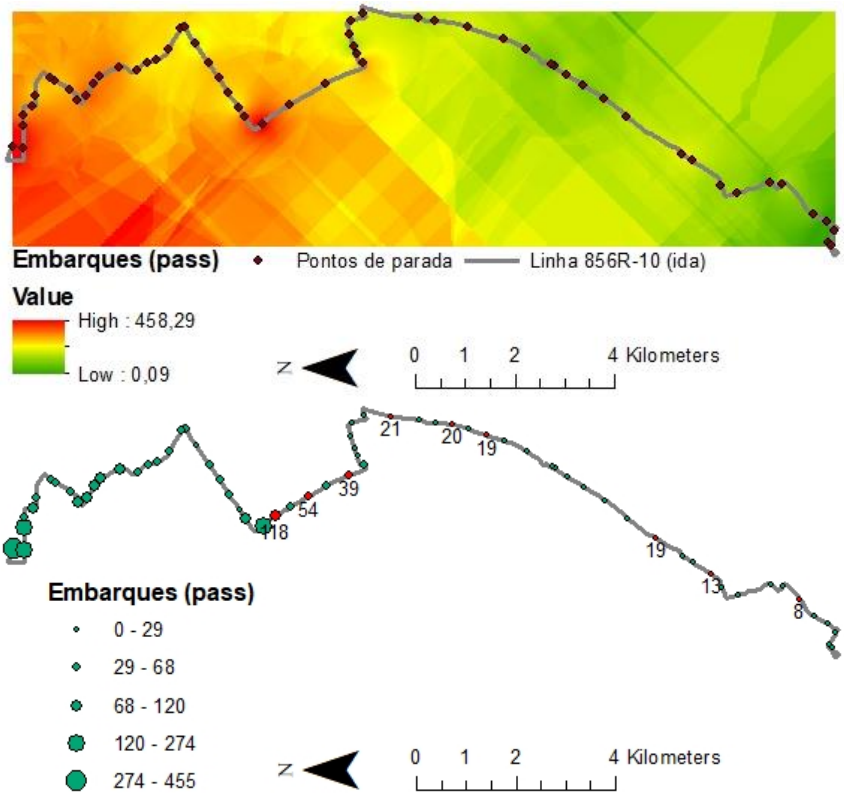
A obtenção de tais resultados sugere que, novamente, o arranjo espacial da linha 856R-10 foi o principal fator interveniente. Enquanto na comparação entre estimativas obtidas por meio de distâncias em rede e euclidianas o traçado relativamente simples do itinerário tornou as duas distâncias praticamente iguais, na presente análise, a orientação quase bem definida da linha é a responsável por aproximar os semivariogramas omnidirecionais daqueles calculados

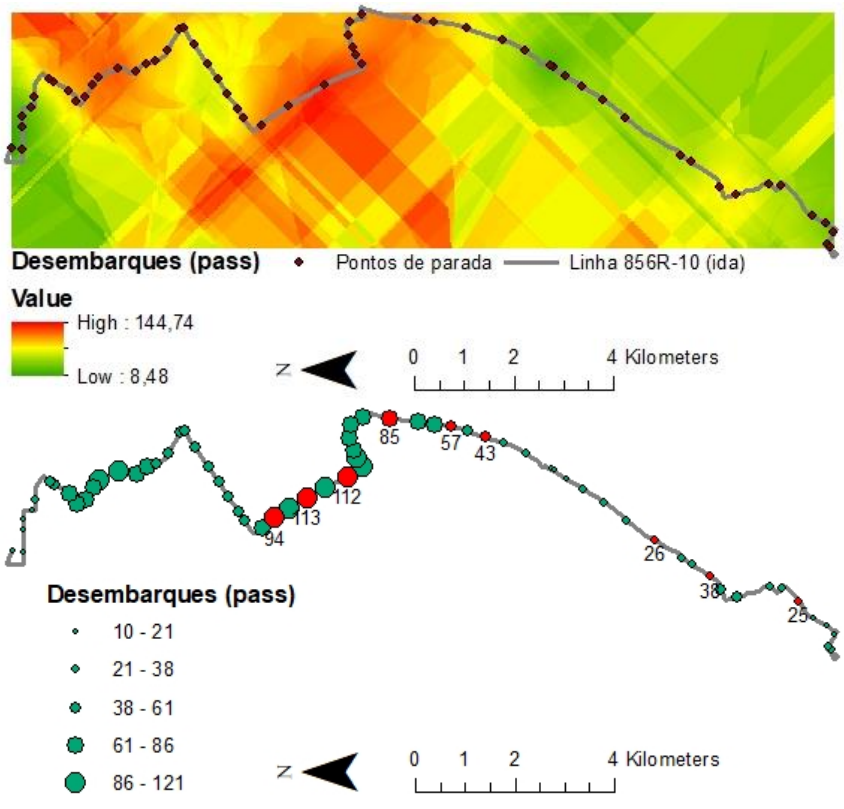
na direção principal. Logo, os pares incluídos no cômputo das semivariâncias médias provavelmente foram os mesmos em ambos os casos, o que levou a parâmetros similares para os modelos teóricos das duas condições e, consequentemente, a estimativas equivalentes.

## 4.3. Superfície de Krigagem

Conforme definido na Seção 3.2.7, 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 12 exemplifica esse procedimento ao mostrar, em b) e d), respectivamente, o número de Embarques e Desembarques atribuído a pontos não amostrados (em vermelho), coletados a partir das superfícies contínuas a) e c), geradas pela Krigagem Ordinária com distâncias euclidianas a partir do semivariograma exponencial, para Embarques, e gaussiano, para Desembarques.

Figura 12 – Estimativa de a) e b) Embarques; e c) e d) Desembarques em pontos não amostrados






c)

d)

Fonte: Elaboração própria

# 5 CONCLUSÕES E RECOMENDAÇÕES PARA TRABALHOS FUTUROS

A pesquisa sobe/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 visitação das linhas de ônibus. Dessa forma, o presente trabalho utilizou a 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, até o presente momento, 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 na etapa de validação cruzada, 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 sobe/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.

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. Assim, responde-se à principal pergunta que norteou este trabalho, descrita no Capítulo 1: *É possível, por meio das técnicas de interpolação geoestatísticas, alcançar estimativas suficientemente precisas do número de Embarques e*

*Desembarques por ponto de parada e Carregamento nos trechos de linha, de forma a tornar desnecessária a realização da pesquisa sobe/desce em toda a rede de TP?*

Retomando o primeiro objetivo específico listado na Seção 1.2, a saber: *propiciar subsídios à não necessidade de realização da pesquisa sobe/desce em todo o itinerário das linhas de ônibus e/ou todas as linhas de determinada rede de TP*, 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 e testes de hipóteses revelaram que, mesmo requerendo um número mínimo de informações para a realização da modelagem geoestatística, a Krigagem Ordinária consegue alcançar ótimas estimativas.

Essa afirmação ganha maiores proporções em se tratando, especialmente, de Carregamento: devido ao fato de ser a mais contínua 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 -5% a 10% 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, consequentemente, 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 Seção 1.3, essa condição ainda não havia sido verificada em trabalhos anteriores (GOMES et al., 2016; LINDNER; 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.

É importante reconhecer, ainda, que tanto os resultados encontrados para distâncias euclidianas quanto para distâncias em rede foram igualmente satisfatórios. A exemplo dos trabalhos citados no capítulo de revisão bibliográfica (SELBY; KOCKELMAN, 2013; ZHANG; WANG, 2014), a utilização de distâncias ao longo da linha de ônibus pouco contribuiu para a melhoria das estimativas. Isso se deve, provavelmente, ao fato da linha 856R-10 ter um arranjo espacial simples, o que torna as distâncias em rede próximas das distâncias euclidianas. Porém, tendo em vista que o traçado dessa linha dificilmente reflete o itinerário

“médio” da maioria das linhas de ônibus existentes na cidade de São Paulo, tais resultados não podem ser generalizados.

A conveniência de se aplicar a KO com distâncias em rede em detrimento das distâncias euclidianas está, dessa forma, condicionada à quantidade de pontos a serem incluídos na modelagem geoestatística. De maneira semelhante ao VMDA, uma quantidade expressiva de dados espaciais demandaria elevados recursos computacionais para a geração das matrizes de distâncias ao longo da linha e, tendo em vista a não ocorrência de diferenças significativas entre os dois tipos de estimativas, o analista de transporte poderia optar, nesse caso, por não utilizar essa abordagem.

A comparação entre as estimativas provenientes do semivariograma omnidirecional e do semivariograma na direção principal (azimute 162°) também não acusou nenhuma diferença significativa. Cabe relembrar que tal procedimento foi realizado a fim de se obter respostas ao seguinte questionamento: se fosse possível calcular semivariogramas em uma direção específica e considerando distâncias em rede, os valores interpolados resultantes dessa modelagem seriam próximos daqueles provenientes do semivariograma direcionado tradicional, com distâncias euclidianas? Novamente, o traçado da linha 856R-10 pode ter sido fator determinante para que não houvesse distinção apreciável entre os dois métodos. Devido à sua orientação quase bem definida, os pares encontrados a cada passo para o cálculo das semivariâncias no azimute principal são praticamente os mesmos que aparecem no semivariograma omnidirecional.

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 estudos futuros, o desempenho da Krigagem Ordinária ao da Krigagem Universal, cuja introdução de variáveis independentes à estrutura do estimador pode levar a resultados ainda melhores. No caso da linha 856R, espera-se que tal aprimoramento beneficie, principalmente, o desempenho das estimativas de Embarques e Desembarques. Porém, a Krigagem Universal provavelmente será uma ferramenta mais adequada que a KO quando forem analisadas outras rotas que não apresentam uma estrutura espacial tão aparente em seu perfil de movimentação de passageiros quanto a selecionada para o presente estudo.

No âmbito da Krigagem Universal, será possível realizar duas novas análises, a saber: 1) extrapolação dos dados da pesquisa sobe/desce para linhas adjacentes, baseado nas informações de trechos que foram visitados pelo referido levantamento; 2) validação

propriamente dita, isto é, avaliar o potencial de estimação da krigagem mediante um cenário simulado em que apenas determinados trechos de linha foram visitados por ocasião da pesquisa sobe/desce. Para o segundo caso, 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.

Ainda, ressalta-se que a contribuição da aplicação da Krigagem Ordinária às variáveis da pesquisa sobe/desce se dá no âmbito operacional, ou seja, no planejamento de curto prazo, o que converge com as constantes mudanças na oferta/demanda do transporte público por ônibus. A Krigagem Universal, por sua vez, permitirá o fornecimento de subsídios ao planejamento de longo prazo, uma vez que a estrutura de variáveis explicativas desse estimador possibilita a análise de diferentes cenários hipotéticos futuros.

A ocorrência de tendências em Embarques, Desembarques e Carregamento, que pode ter sido fator contribuinte para a obtenção de erros extremos e considerável porcentagem de efeito pepita, no caso das duas primeiras variáveis, e de um elevado parâmetro de alcance em Carregamento, sugere, além da Krigagem Universal, a utilização da Krigagem com Deriva Externa. Esse estimador também é recomendado à modelagem de variáveis que possuam altas aleatoriedades na vizinhança próxima ou variações de longo alcance, provavelmente provenientes da influência de outras variáveis.

Além disso, tendo em vista as respostas encontradas para os objetivos específicos, outras recomendações para trabalhos futuros podem ser descritas. Umas das principais seria estender a análise empreendida no presente estudo a um número considerável de linhas, a fim de refinar e fortalecer os resultados e conclusões alcançados. Além disso, o mesmo método poderia ser aplicado às variáveis de Embarques, Desembarques e Carregamento em seu estado bruto, sem a transformação de Box-Cox, no intuito de verificar se há, realmente, alguma melhoria nas estimativas quando a modelagem geoestatística é baseada nos dados transformados. Por ser um artifício comumente usado, os autores optaram por incluí-lo no método, porém, nenhum dos trabalhos citados esclarece, de fato, se ocorre diferenças significativas quando as transformações não são empregadas.

Ademais, recomenda-se verificar a influência do tamanho da grade no cálculo das estimativas de krigagem provenientes da superfície de interpolação. Apesar de se ter considerado aceitável a resolução adotada no presente estudo, analisar potencial ganho ou perda no poder preditivo da técnica resultante de variações no tamanho da grade pode contribuir para a obtenção de um desempenho otimizado dos resultados, tanto com relação às métricas de

aderência quanto aos recursos computacionais e de tempo de processamento exigidos pela modelagem.

Uma quarta sugestão se refere aos questionamentos acerca da utilização de distâncias em rede e semivariogramas na direção principal. Trata-se da substituição do semivariograma tradicional pelo designado semivariograma regularizado, no qual adota-se, ao invés das distâncias euclidianas e em rede entre pares, o valor equivalente das mesmas projetado em um eixo vertical que aponta para o norte geográfico. Esse método eliminaria qualquer influência do tipo de distância ou azimute considerado no cálculo do semivariograma experimental.

Finalmente, sugere-se o uso de simulação geoestatística para variáveis de demanda por transporte público como as utilizadas neste trabalho. Seria interessante testar o uso de Simulação Sequencial Guassiana, para se ter um intervalo de confiança dos valores estimados a partir das simulações. O planejador de transportes, teria, então, informações relativas aos piores cenários por coordenadas geográficas (menores números de usuários) e melhores cenários (maiores números de usuários). Tais cenários poderiam ser eficientes para planejamento da oferta considerando horários de pico e entre picos.

# REFERÊNCIAS


BAGCHI, M.; WHITE, P. R. The potential of public transport smart card data. **Transport Policy**, v. 12, n. 5, p. 464–474, 2005.

BARBOSA, H. M.; BRAGA, C. M. P.; COELHO, J. A. P.; ARRUDA, T.; JOTA, F. G. **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. **Anais**...2006

BARTLETT, M. S. The Use of Transformations. **Biometrics**, v. 3, n. 1, p. 39–52, 1947.

BAYRAKTAR, H.; TURALIOGLU, F. S. A Kriging-based approach for locating a sampling site in the assessment of air quality. **Stochastic Environmental Research and Risk Assessment**, v. 19, n. 4, p. 301–305, 2005.

BENEDICT, S.; ERIC, A.; JOSEPH, M. Linear Spatial Interpolation and Analysis of Annual Average Daily Traffic Data. **Journal of Computing in Civil Engineering**, v. 29, n. 1, p. 4014022, 1 jan. 2015.

BOX, G. E. P.; COX, D. R. An Analysis of Transformations. **Journal of the Royal Statistical Society. Series B (Methodological)**, v. 26, n. 2, p. 211–252, 1964.

BUNDALA, D.; BERGENHEIM, W.; METZ, M. v.net.allpairs - Computes the shortest path between all pairs of nodes in the network. GRASS GIS code, 2014. Disponível em: < https://trac.osgeo.org/grass/browser/grass/branches/releasebranch_7_2/vector/v.net.allpairs>

BURGESS, T. M.; WEBSTER, R. Optimal Interpolation and Isarithmic Mapping of Soil Properties. **Journal of Soil Science**, v. 31, n. 2, p. 333–341, 28 jul. 2006.

CEDER, A. **Public transit planning and operation: modeling, practice and behavior**. CRC press, 2007.

CHI, G.; ZHENG, Y. Estimating Transport Footprint along Highways at Local Levels: A Combination of Network Analysis and Kriging Methods. **International Journal of Sustainable Transportation**, v. 7, n. 3, p. 261–273, 2013.

CRESSIE, N. Fitting variogram models by weighted least squares. **Journal of the International Association for Mathematical Geology**, v. 17, n. 5, p. 563–586, 1985.

CRESSIE, N. **Statistics for spatial data**. John Wiley & Sons, Inc., 1993.

CRESSIE, N.; HAWKINS, D. M. Robust estimation of the variogram. **I. Journal of the International Association for Mathematical Geology**, vol. 12, n. 2, p. 115-125, 1958.

EOM, J. K; PARK, M. S.; HEO, T.; HUNTSINGER, L. F. 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, 2006.

ESRI. Geostatistical Analyst Tutorial. Environmental Systems Research Institute, 2010. Disponível em: < http://help.arcgis.com/en/arcgisdesktop/10.0/pdf/geostatistical-analyst-tutorial.pdf> Acesso em: mai. 2018.

FERRAZ, A. C. P.; TORRES, I. G. E. **Transporte público urbano**. 2. ed. RiMa Editora, 2004.

GOMES, M. M.; PITOMBO, C. S.; PIRDAVANI, A.; BRIJS, T.. Geostatistical approach to estimate car occupant fatalities in traffic accidents. **Revista Brasileira de Cartografia**, v. 70, n. 4, p. 1231–1256, 2018.

GOMES, V. A.; PITOMBO, C. S.; ROCHA, S. S.; SALGUEIRO, A. R. Kriging Geostatistical Methods for Travel Mode Choice: A Spatial Data Analysis to Travel Demand Forecasting. **Open Journal of Statistics**, v. 06, n. 03, p. 514–527, 2016.

GOOVAERTS, P. **Geostatistics for Natural Resources and Evaluation**. Oxford University Press, 1997.

GOOVAERTS, P. Kriging and Semivariogram Deconvolution in the Presence of Irregular Geographical Units. **Mathematical geology**, v. 40, n. 1, p. 101–128, 2008.

GOOVAERTS, P. Combining area-based and individual-level data in the geostatistical mapping of late-stage cancer incidence. **Spatial and Spatio-temporal Epidemiology**, v. 1, n. 1, p. 61–71, 2009.

GOOVAERTS, P. Geostatistical analysis of health data with different levels of spatial aggregation. **Spatial and Spatio-Temporal Epidemiology**, v. 3, n. 1, p. 83–92, 11 abr. 2012.

GUNDOGDU, I. B.; FACULTY, E. Risk governance for traffic accidents by Geostatistical Analyst methods. **International Journal of Research in Engineering and Science**, v. 2, n. 9, p. 35–40, 2014.

HIBBERT, J. D.; LIESE, A. D.; LAWSON, A.; PORTER, D. E.; PUETT, R. C.; STANDIFORD, D.; LIU, L.; DABELEA, D. Evaluating geographic imputation approaches for zip code level data: an application to a study of pediatric diabetes. **International Journal of Health Geographics**, v. 8, p. 54, 8 out. 2009.

HOLLANDER, Y.; LIU, R. The principles of calibrating traffic microsimulation models. **Transportation**, v. 35, n. 3, p. 347–362, 2008.

IBGE. São Paulo. IBGE Cidades, Instituto Brasileiro de Geografia e Estatística, 2018. Disponível em: <https://cidades.ibge.gov.br/brasil/sp/sao-paulo/panorama>. Acesso em: mai. 2018.

IBM. IBM SPSS Statistics 24 Core System User's Guide. International Business Machines, 2016. 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.

JOURNEL, A. G.; HUIJBREGTS, C. J. **Mining geostatistics**. Academic press, 1978.

KLATKO, T. J.; SAEED, T. U.; VOLOVSKI, M.; LABI, S.; FRICKER, J. D.; SINHA, K. C. Addressing the Local-Road VMT Estimation Problem Using Spatial Interpolation Techniques. **Journal of Transportation Engineering, Part A: Systems**, v. 143, n. 8, p. 4017038, 1 ago. 2017.

KRIGE, D. G. A statistical approach to some basic mine valuation problems on the Witwatersrand. **Journal of the Southern African Institute of Mining and Metallurgy**, v. 52, n. 6, p. 119–139, 1951.

LANDIM, P. M. B. Sobre geoestatística e mapas. **Terrae Didatica**, v. 2, n. 1, p. 19–33, 2006.

LASLETT, G. M.; McBRATNEY, A. B.; PAHL, P. J.; HUTCHINSON, M. F. Comparison of several spatial prediction methods for soil pH. **Journal of Soil Science**, v. 38, n. 2, p. 325–341, 28 jul. 2006.

LINDNER, A.; PITOMBO, C. S.; ROCHA, S. S.; QUINTANILHA, J. A. Estimation of transit trip production using Factorial Kriging with External Drift: an aggregated data case study. **Geo-spatial Information Science**, v. 19, n. 4, p. 245–254, 1 out. 2016.

LINDNER, A.; PITOMBO, C. S. A Conjoint Approach of Spatial Statistics and a Traditional Method for Travel Mode Choice Issues. **Journal of Geovisualization and Spatial Analysis**, v. 2, n. 1, p. 1, 2017.

MAJUMDAR, A.; NOLAND, R. B.; OCHIENG, W. Y. A spatial and temporal analysis of safety-belt usage and safety-belt laws. **Accident Analysis & Prevention**, v. 36, n. 4, p. 551–560, 2004.

MANEPALLI, U. R. R.; BHAM, G. H. **Crash prediction: evaluation of empirical Bayes and kriging Methods**. 3rd International Conference on Road Safety and Simulation. **Anais**...2011

MATHERON, G. Principles of geostatistics. **Economic Geology**, v. 58, n. 8, p. 1246–1266, dez. 1963.

MATHERON, G. The Theory of Regionalized Variables and Its Applications. Les Cahiers du Centre de Morphologie Mathematique in Fontainebleu, Paris, 1971.

MEYER, M. D.; INSTITUTE OF TRANSPORTATION ENGINEERS. Introduction to Transportation Planning. **Transportation Planning Handbook**, v. 3, p. 12, 2016.

MILLARD, S. P. **EnvStats**. Springer eBooks, v. 2 2013

MORAN, P. A. P. The interpretation of statistical maps. **Journal of the Royal Statistical Society. Series B (Methodological)**, v. 10, n. 2, p. 243–251, 1948.

OLEA, R. A. A six-step practical approach to semivariogram modeling. **Stochastic Environmental Research and Risk Assessment**, v. 20, n. 5, p. 307–318, 2006.

OLIVER, M. A.; WEBSTER, R. Combining Nested and Linear Sampling for Determining the

Scale and Form of Spatial Variation of Regionalized Variables. **Geographical Analysis**, v. 18, n. 3, p. 227–242, 3 set. 2010.

OLIVER, M. A.; WEBSTER, R. **Basic steps in geostatistics: the variogram and kriging**. Springer, 2015.

PARADIS, E.; CLAUDE, J.; STRIMMER, K. APE: analyses of phylogenetics and evolution i n R language. **Bioinformatics,** v. 20, p. 289-290, 2004.

PITOMBO, C. S.; SALGUEIRO, A. R.; COSTA, A. S. G.; ISLER, C. A. A two-step method for mode choice estimation with socioeconomic and spatial information. **Spatial Statistics**, v. 11, p. 45–64, 2015.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2018. URL https://www.R-project.org/.

RIBEIRO JR., P. J.; DIGGLE, P. J. geoR: Analysis of Geostatistical Data. R package version 1.7-5.2. 2016. https://CRAN.R-project.org/package=geoR

ROCHA, S. S.; LINDNER, A.; PITOMBO, C. S. Proposal of a geostatistical procedure for tra nsportation planning field, **Boletim de Ciências Geodésicas**, Scielo, 2017.

RODRIGUE, J. P.; COMTOIS, C.; SLACK, B. **The geography of transport systems**. 4. ed. New York: Routledge, 2017.

SÃO PAULO. **Plano de Mobilidade de São Paulo**. Prefeitura do Município de São Paulo, Secretaria Municipal de Transportes, 2015. Disponível em: < http://www.prefeitura.sp.gov.br/cidade/secretarias/upload/chamadas/planmobsp_v072__1455 546429.pdf>. Acesso em: mai. 2018.

SELBY, B.; KOCKELMAN, K. M. Spatial prediction of traffic levels in unmeasured locations: Applications of universal kriging and geographically weighted regression. **Journal of Transport Geography**, v. 29, p. 24–32, 2013.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, v. 52, n. 3–4, p. 591–611, 1 dez. 1965.

SPTRANS. Investimentos e Responsabilidades. Prefeitura de São Paulo, Edital 080715 (Anexo VI), 2015. Disponível em: <http://www.sptrans.com.br/downloads/Editais_080715/ESTRUTURAL/Estrutural-AnexoVI.pdf>. Acesso em: mar. 2018

SPTRANS. Dados SPTrans 2017. São Paulo Transporte S.A., 2018.

SUBHADIP, B.; CHAKRABORTY, S.; CHANDRA, S.; GHOSH, I. Kriging-Based Approach for Estimation of Vehicular Speed and Passenger Car Units on an Urban Arterial. **Journal of Transportation Engineering, Part A: Systems**, v. 143, n. 3, p. 4016013, 1 mar. 2017.

VER HOEF, J. M. Kriging models for linear networks and non-Euclidean distances: Cautions

and solutions. **Methods in Ecology and Evolution**, v. 9, n. 6, p. 1600–1613, 31 jan. 2018.

VOGEL, R. M. The Probability Plot Correlation Coefficient Test for the Normal, Lognormal, and Gumbel Distributional Hypotheses. **Water Resources Research**, v. 22, n. 4, p. 587–590, abr. 1986.

VUCHIC, V. R. **Urban Transit: Operations, Planning, and Economics**. John Wiley & Sons, 2005.

WANG, X.; KOCKELMAN, K. Forecasting Network Data. **Transportation Research Record: Journal of the Transportation Research Board**, v. 2105, p. 100–108, 6 out. 2009.

YAMAMOTO, J. K.; LANDIM, P. M. B. **Geoestatística: conceitos e aplicações**. Oficina de textos, 2015.

ZHANG, D.; WANG, X. C. Transit ridership estimation with network Kriging: A case study of Second Avenue Subway, NYC. **Journal of Transport Geography**, v. 41, p. 107–115, 2014.

ZIMMERMAN, D. A.; MARSILY, G.; GOTWAY, C. A.; MARIETTA, M. G.; AXNESS, C. L.; BEAUHEIM, R. L.; BRAS, R. L.; CARRERA, J.; DAGAN, G.; DAVIES, P. B. A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow. **Water Resources Research**, v. 34, n. 6, p. 1373–1413, 1998.

# APÊNDICE

Tabela 10 - Testes de hipóteses não-paramétricos aplicados à comparação entre os valores reais e as estimativas da Krigagem Ordinária omnidirecional com distâncias euclidianas

| Variável | Modelo teórico | Teste | Sig. | Conclusão |
|---|---|---|---|---|
| Embarques | Exponencial | Mediana | 0,708 | Não há diferenças significativas |
| | | Mann-Whitney | 0,998 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,981 | Não há diferenças significativas |
| | Gaussiano | Mediana | 0,708 | Não há diferenças significativas |
| | | Mann-Whitney | 0,930 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,783 | Não há diferenças significativas |
| | Esférico | Mediana | 0,708 | Não há diferenças significativas |
| | | Mann-Whitney | 0,975 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,910 | Não há diferenças significativas |
| Desembarques | Exponencial | Mediana | 0,708 | Não há diferenças significativas |
| | | Mann-Whitney | 0,966 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,160 | Não há diferenças significativas |
| | Gaussiano | Mediana | 0,708 | Não há diferenças significativas |
| | | Mann-Whitney | 0,916 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,103 | Não há diferenças significativas |
| | Esférico | Mediana | 0,708 | Não há diferenças significativas |
| | | Mann-Whitney | 0,930 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,103 | Não há diferenças significativas |
| Carregamento | Exponencial | Mediana | 0,571 | Não há diferenças significativas |
| | | Mann-Whitney | 0,935 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,999 | Não há diferenças significativas |
| | Gaussiano | Mediana | 0,850 | Não há diferenças significativas |
| | | Mann-Whitney | 1,000 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,617 | Não há diferenças significativas |
| | Esférico | Mediana | 0,571 | Não há diferenças significativas |
| | | Mann-Whitney | 0,940 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,999 | Não há diferenças significativas |

Fonte: Elaboração própria

Tabela 11 - Testes de hipóteses não-paramétricos aplicados à comparação entre os valores reais e as estimativas da Krigagem Ordinária omnidirecional com distâncias em rede

| Variável | Modelo teórico | Teste | Sig. | Conclusão |
|---|---|---|---|---|
| Embarques | Exponencial | Mediana | 1,000 | Não há diferenças significativas |
| | | Mann-Whitney | 0,984 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,910 | Não há diferenças significativas |
| | Gaussiano | Mediana | 0,708 | Não há diferenças significativas |
| | | Mann-Whitney | 0,921 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,628 | Não há diferenças significativas |
| | Esférico | Mediana | 0,708 | Não há diferenças significativas |
| | | Mann-Whitney | 0,953 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,783 | Não há diferenças significativas |
| Desembarques | Exponencial | Mediana | 0,708 | Não há diferenças significativas |
| | | Mann-Whitney | 0,998 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,344 | Não há diferenças significativas |
| | Gaussiano | Mediana | 1,000 | Não há diferenças significativas |
| | | Mann-Whitney | 0,962 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,344 | Não há diferenças significativas |
| | Esférico | Mediana | 1,000 | Não há diferenças significativas |
| | | Mann-Whitney | 0,980 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,476 | Não há diferenças significativas |
| Carregamento | Exponencial | Mediana | 0,850 | Não há diferenças significativas |
| | | Mann-Whitney | 0,991 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 1,000 | Não há diferenças significativas |
| | Gaussiano | Mediana | 0,345 | Não há diferenças significativas |
| | | Mann-Whitney | 0,884 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 0,774 | Não há diferenças significativas |
| | Esférico | Mediana | 0,850 | Não há diferenças significativas |
| | | Mann-Whitney | 0,963 | Não há diferenças significativas |
| | | Kolmogorov-Smirnov | 1,000 | Não há diferenças significativas |

Fonte: Elaboração própria

Figura 13 – Semivariogramas empíricos de Embarques (azimute entre parênteses na legenda do eixo das ordenadas)

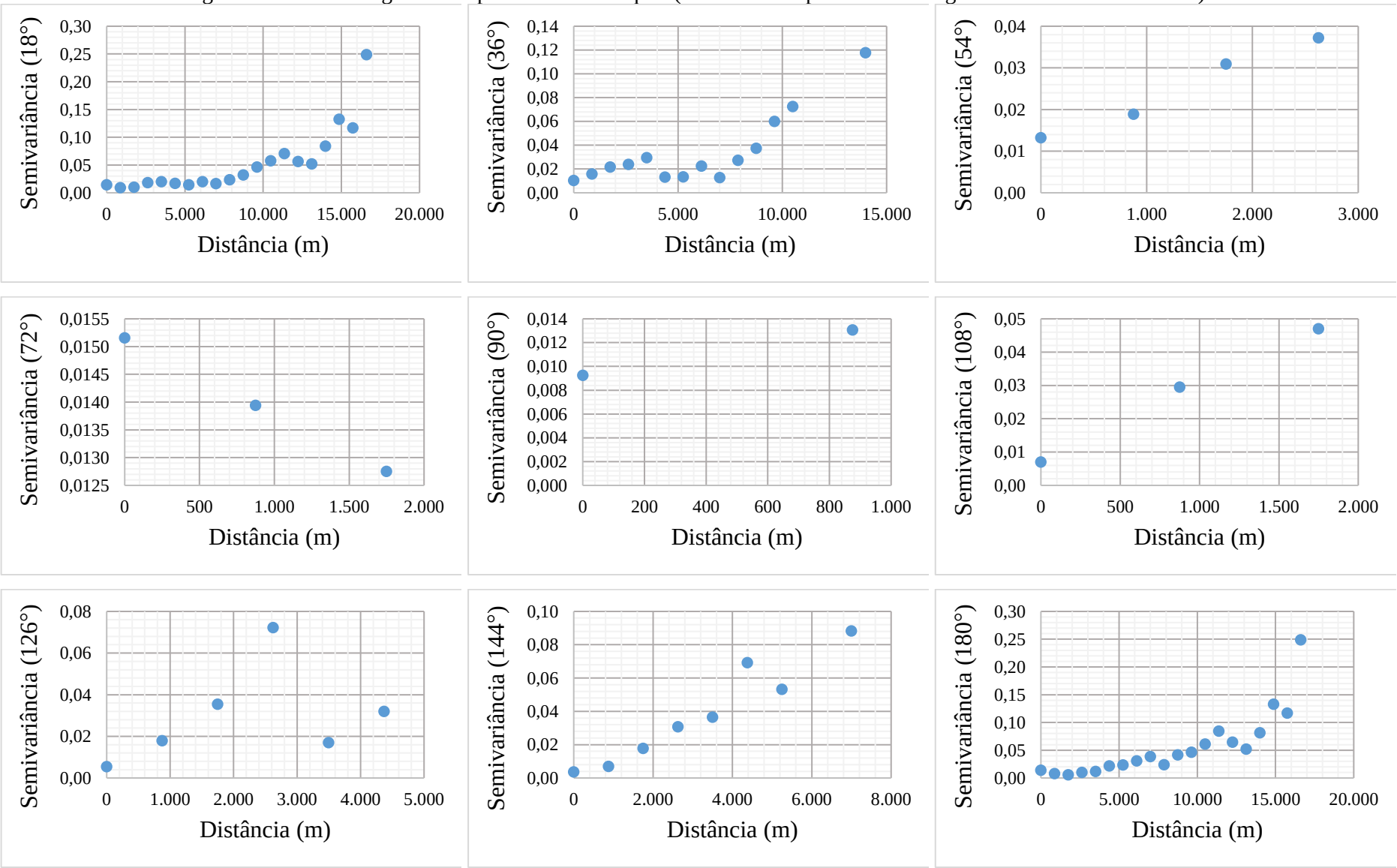


Figura 14 - Semivariogramas empíricos de Desembarques (azimute entre parênteses na legenda do eixo das ordenadas)

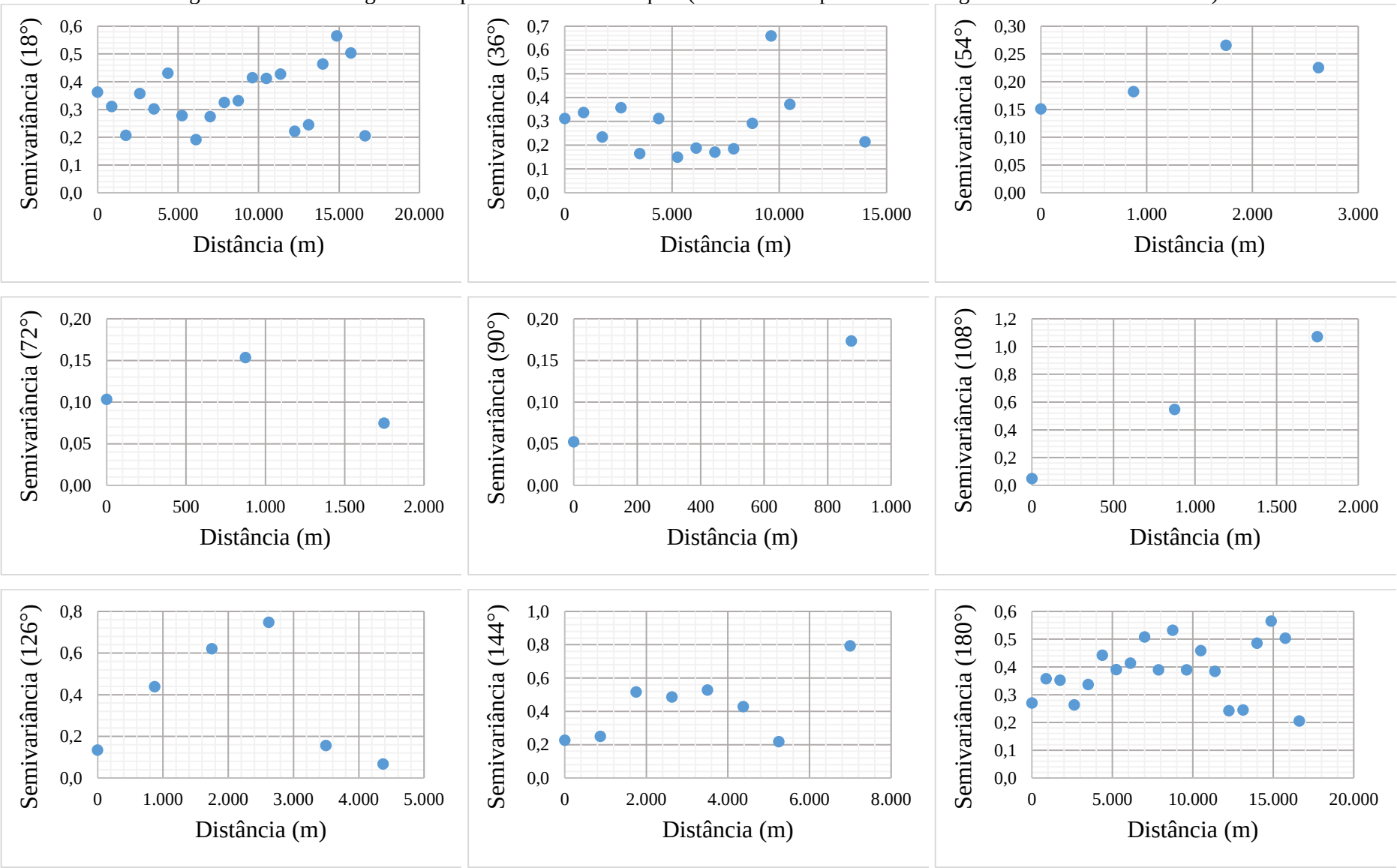


Figura 15 - Semivariogramas empíricos de Carregamento (azimute entre parênteses na legenda do eixo das ordenadas)

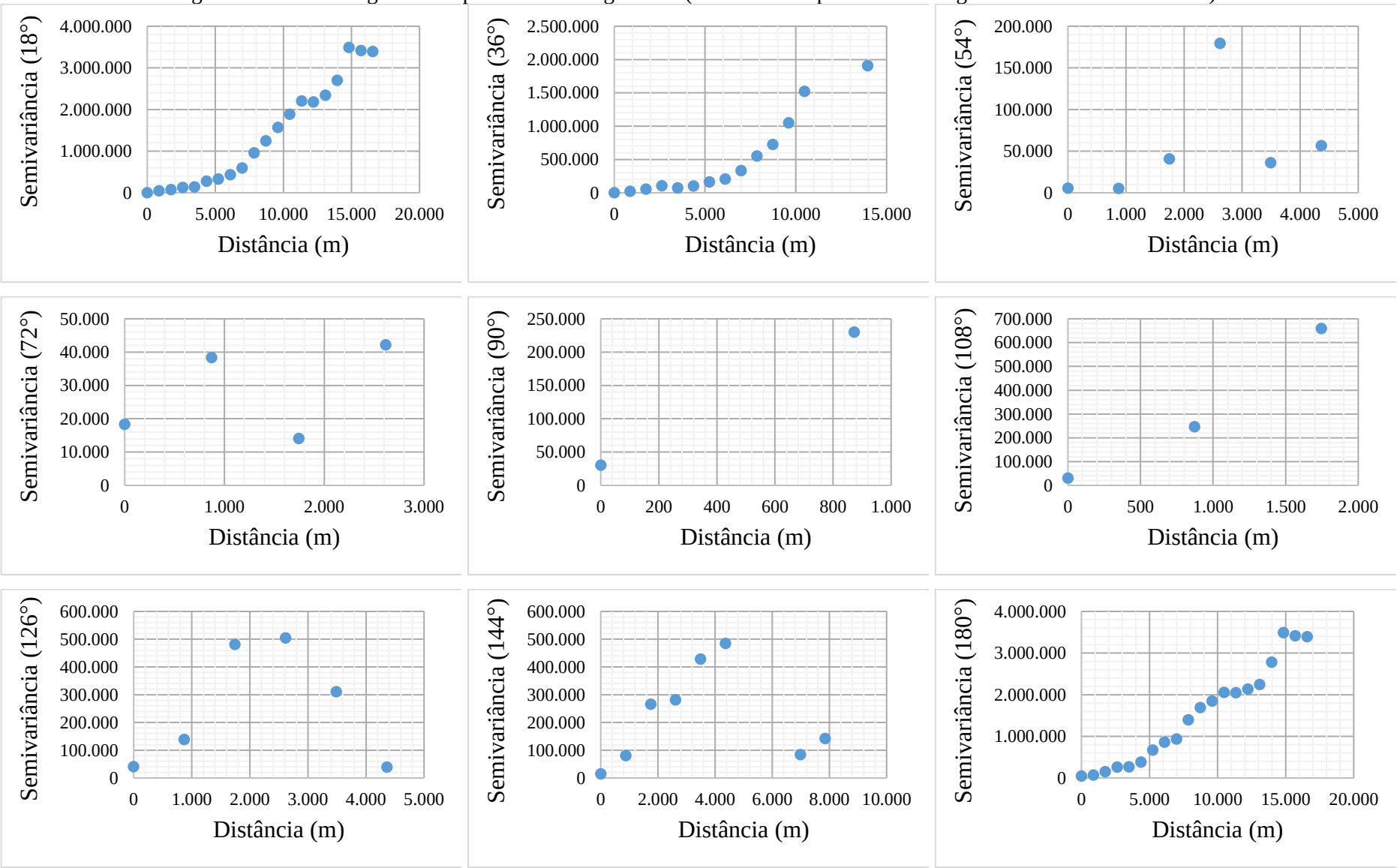
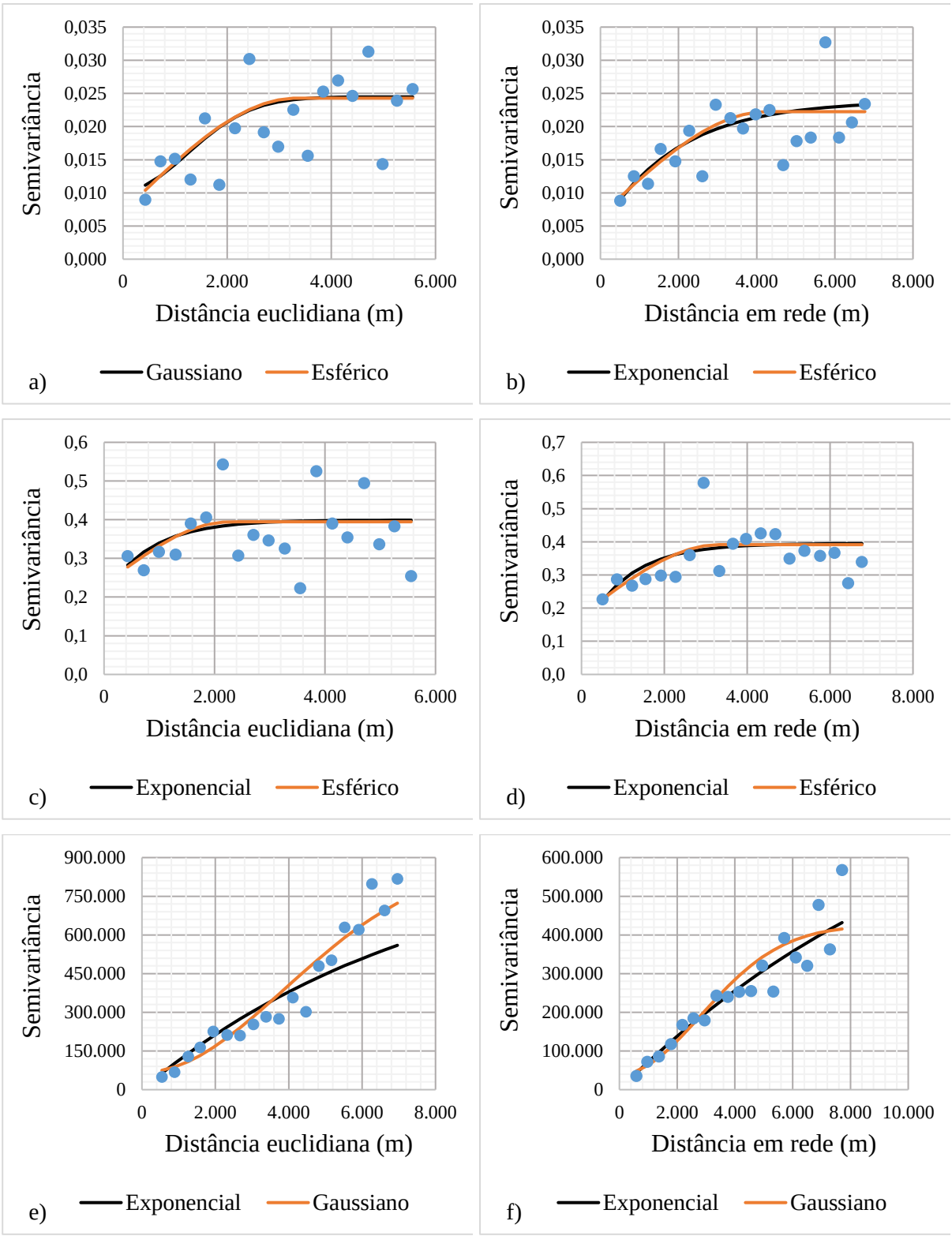



Figura 16 – Modelos teóricos para os semivariogramas de a) Embarques com distâncias euclidianas; b) Embarques com distâncias em rede; c) Desembarques com distâncias euclidianas; d) Desembarques com distâncias em rede; e) Carregamento com distâncias euclidianas; e f) Carregamento com distâncias em rede