ALINE SANTOS DAMASCENA

**Modelo espaço-temporal para a predição da concentração de material
particulado fino na Região Metropolitana de São Paulo**

Tese apresentada à Faculdade de Medicina da
Universidade de São Paulo para obtenção do
título de Doutor em Ciências

Programa: Patologia
Orientador: Prof. Dr. Paulo Hilario
Nascimento Saldiva

Versão corrigida. Resolução CoPGr 6018, de 13 de outubro de 2011.
(Versão original disponível na Biblioteca da FMUSP)

São Paulo
2020

Dados Internacionais de Catalogação na Publicação (CIP)

Preparada pela Biblioteca da
Faculdade de Medicina da Universidade de São Paulo

©reprodução autorizada pelo autor

Damascena, Aline Santos

Modelo espaço-temporal para a predição da
concentração de material particulado fino na Região
Metropolitana de São Paulo / Aline Santos
Damascena. -- São Paulo, 2020.

Tese(doutorado)--Faculdade de Medicina da
Universidade de São Paulo.

Programa de Patologia.

Orientador: Paulo Hilario Nascimento Saldiva.

Descritores: 1.Poluição atmosférica 2.Material
particulado 3.Modelos de qualidade do ar
4.Tecnologia de sensoriamento remoto 5.Imagens por
satélite 6.Modelos de predição 7.Modelos estatísticos
8.Modelos lineares 9.Exposição ambiental

USP/FM/DBD-247/20

Responsável: Erinalva da Conceição Batista, CRB-8 6755

Agradecimentos

Gostaria de agradecer à **Universidade de São Paulo** e a todos que trabalham para torná-la uma instituição de excelência. Graças a vocês, a Universidade conta com uma estrutura que possibilitou que eu finalizasse esta tese à distância, na segurança da minha casa, em plena pandemia de COVID-19. Muito obrigada!

Agradeço ao **Professor Paulo Saldiva** pela oportunidade e confiança. Também sou grata aos **Professores Francisco Marcelo Monteiro Rocha, Carmen Saldiva e Nelson Tanaka**, do Instituto de Matemática e Estatística (IME), que foram meus guias nessa jornada.

Da mesma forma agradeço à **Professora Leila Droprinchinski Martins** da Universidade Tecnológica Federal do Paraná, ao **Professor Gláuber Guimarães Cirino da Silva** da Universidade Federal do Pará e ao **Professor Julio da Motta Singer** do IME, pela leitura cuidadosa do documento da tese e pelas contribuições valiosas feitas durante a banca de avaliação.

O projeto multidisciplinar que esta tese pertence me deu a oportunidade de interagir e aprender com muitos professores e colegas, o que com certeza contribuiu imensamente para a minha pesquisa.

Agradeço ao **Professor Paulo Polito** do Instituto Oceanográfico, por me aceitar como aluna ouvinte na disciplina IOF0255 - Oceanografia por Satélite, onde tive o primeiro contato com os conceitos de sensoriamento remoto por satélites. Também agradeço aos **Professores Alexandre Correia** do Instituto de Física, **Regina Maura de Miranda** da Escola de Ciências, Artes e Humanidades, **Maria de Fátima Andrade** do Instituto de Astronomia, Geofísica e Ciências Atmosféricas (IAG), **Lígia Vizeu Barrozo** da Faculdade de Filosofia, Letras e Ciências Humanas (FFLCH), **Fábio Lopes** do Instituto de Pesquisas Energéticas e Nucleares (IPEN) e **Débora de Souza Alvim** do Instituto Nacional de Pesquisas Espaciais (INPE), pela disponibilidade em um momento em que eu estava cheia de dúvidas relacionadas a minha pesquisa.

Em especial, sou grata à **Professora Márcia Akemi Yamasoe** do IAG, pela parceria e paciência. Foram alguns encontros vasculhando os dados junto comigo na tentativa de entender "Por que não correlaciona?" até que finalmente entendemos!

Meus sinceros agradecimentos ao **Vitor Sousa Martins**, o qual não tive a oportunidade de conhecer pessoalmente, mas desde o primeiro contato via *e-mail* (cheio de dúvidas referentes aos dados de satélite), mostrou-se muito solícito e generoso. Muito obrigada pela

sua disponibilidade e apoio mesmo tendo as demandas do seu próprio doutorado.

Aos novos amigos que fiz no IAG, em especial a **Francisco Daniel Mota Lima, Veronika Brand, Miriam Matias, Caroline Bernardes, Luís Henrique Mendes, Ana Paula Emygdio, Luana Ribeiro, Mário Calderon, Noelia Benavente, Tailine Santos, Jorge Santana e Maciel Sánchez**, pelos inúmeros esclarecimentos relacionados à poluição do ar ou à meteorologia, e principalmente, pelos momentos de alegria e descontração que deixaram essa jornada mais leve.

Aos amigos de longa data **Anelise Guimarães Sabbag, Elivane da Silva Victor, Elizângela Salatiel, Igor André Milhorança, Orlando Yesid, Rafael de Assumpção Sampaio, Verônica Leão Neves, Diego Ferreira Schiezari e Renato Tadashi Izawa** pelo companheirismo, nos bons e maus momentos.

Aos meus pais, **Alírio de Araújo Damascena e Maria de Fátima Lima Santos**, que não mediram esforços para proporcionar a mim e a minha irmã, **Daniele Santos Damascena Pedro**, tudo o que foi necessário para a nossa formação acadêmica. Nada que eu escreva aqui será o suficiente para agradecê-los. Amo vocês!

E por último, mas não menos importante, agradeço ao apoio financeiro da **Coordenação de Aperfeiçoamento de Pessoal de Nível Superior** (CAPES, processo número 001) e da **Fundação de Amparo à Pesquisa do Estado de São Paulo** (FAPESP, processo número 2016/09411-1).

Resumo

Damascena, AS. *Modelo espaço-temporal para a predição da concentração de material particulado fino na Região Metropolitana de São Paulo* [tese]. São Paulo: Faculdade de Medicina da Universidade de São Paulo; 2020.

O padrão espaço-temporal das concentrações do material particulado fino (MP_{2.5}) é um importante fator para prever alguns desfechos de saúde dos habitantes de uma área urbana. As concentrações de MP_{2.5} podem ser obtidas por meio de redes de monitoramento da qualidade do ar, que no geral, possuem um número insuficiente de estações para representar a exposição de toda uma população em uma determinada região. Nesse sentido, do ponto de vista epidemiológico, se faz necessário o desenvolvimento de metodologias que permitam estimar com alta resolução espacial e temporal as concentrações do MP_{2.5} em locais sem estações de monitoramento, de forma que seja possível estudar os desfechos de curto (dias/semanas) e de longo prazo (anos). Nesta tese aplicamos uma metodologia proposta por pesquisadores da Faculdade de Saúde Pública de Harvard para prever as concentrações de MP_{2.5} na Região Metropolitana de São Paulo (RMSP), entre janeiro de 2012 e dezembro de 2017. Além disso, propomos alterações nessa metodologia, com o intuito de considerar nos modelos de predição as duas fontes de correlação (espacial e temporal) inerente aos dados, já que a metodologia proposta por Harvard considera apenas fonte de correlação espacial. Também apresentamos os desvios-padrão dos erros de predição, uma medida da incerteza associada às estimativas obtidas nesses modelos. A metodologia de Harvard tem como base o Modelo Linear Misto (MLM) e a utilização de dados de Profundidade Óptica do Aerossol (POA) obtidos de satélites. A POA tem se mostrado como preditora das concentrações de MP_{2.5} e/ou MP₁₀ medidos na superfície em diferentes regiões do mundo. Com o MLM é possível incorporar nas predições a variação diária da relação POA-MP_{2.5}, ocasionada pelas condições meteorológicas. Pela primeira vez na RMSP, nós utilizamos os dados de POA com alta resolução espacial (1 km²) da versão 6 do algoritmo *Multiangle Implementation of Atmospheric Correction* (MAIAC). Os resultados mostraram que, diferente do que se observa em outras regiões onde essa metodologia já foi utilizada, os dados de POA do MAIAC não foram preditores das concentrações de MP_{2.5} e de MP₁₀ na superfície da RMSP. Verificamos que o modelo de aerossol empregado pelo MAIAC na aquisição dos dados de POA não representa de forma adequada as propriedades do aerossol dessa região. Porém, os modelos de predição propostos que não consideraram esses dados foram capazes de explicar respectivamente, cerca de 60% e 70% da variabilidade das concentrações diurnas e diárias do MP_{2.5} na RMSP. O desempenho dos modelos de predição para as concentrações diárias de MP_{2.5} foi bastante satisfatório e comparável ao desempenho dos modelos de predição desenvolvidos na Suíça, Israel e Cidade do México, que utilizaram os dados de POA do MAIAC. Além disso, os modelos de predição que consideraram as duas fontes de correlação apresentaram poder de predição cerca de 5% maiores do que os modelos que consideraram apenas a correlação

espacial, além de gerarem estimativas mais precisas. Os mapas com a média anual das estimativas das concentrações diurnas e diárias de $MP_{2.5}$ sobre a RMSP foram disponibilizados em um repositório de acesso livre. Essas estimativas podem ser utilizadas como uma aproximação para a exposição à poluição atmosférica dos habitantes da RMSP, contribuindo com os estudos epidemiológicos que visam entender seus impactos na saúde da população.

Descritores: Poluição atmosférica; Material particulado; $MP_{2.5}$; MP_{10} ; Aerossol atmosférico; Profundidade Óptica do Aerossol; Modelos de qualidade do ar; Tecnologia de sensoriamento remoto; Imagens por satélite; Modelos de predição; Modelos estatísticos; Modelos lineares; Exposição ambiental.

Abstract

Damascena, AS. *Spatiotemporal model for predicting fine particulate matter concentration in the Metropolitan Area of São Paulo* [thesis]. São Paulo: “Faculdade de Medicina, Universidade de São Paulo”; 2020.

The spatiotemporal pattern of fine particulate matter (PM_{2.5}) concentrations is an important factor in predicting some health outcomes of inhabitants in urban areas. PM_{2.5} concentrations are measured in air quality monitoring networks, which, in general, are sparse and do not represent the exposure of an entire population in a given region. From an epidemiological point of view, it is necessary to develop methodologies that allow us to estimate PM_{2.5} concentrations with high spatial and temporal resolution in places without monitoring stations, so that it is possible to study the short-term (days/weeks) and long term (years) outcomes. In this thesis, we applied a methodology proposed by researchers from the Harvard School of Public Health to predict the concentrations of PM_{2.5} in the Metropolitan Area of São Paulo (MASP), in the 2012-2017 period. In addition, we proposed changes in that methodology, in order to consider in the prediction models the two sources of correlation (spatial and temporal) inherent to the data, since the methodology proposed by Harvard considers only the spatial correlation. We also presented the standard deviations of the prediction errors associated with the estimates obtained in those models. The Harvard methodology is based on the Linear Mixed Model (LMM) and on the use of Aerosol Optical Depth (AOD) data obtained from satellites. AOD data have been shown to be a predictor of surface PM_{2.5} and/or PM₁₀ concentrations in different regions of the world. With LMM it is possible to incorporate in the predictions the daily variation of AOD-PM_{2.5} relationship, caused by the variation of meteorological conditions. For the first time in the MASP, we used AOD data with high spatial resolution (1 km²) from version 6 of the Multiangle Implementation of Atmospheric Correction (MAIAC) algorithm. Our results showed that, unlike what was observed in other regions where Harvard methodology has already been applied, AOD data from MAIAC were not predictive of surface PM_{2.5} and PM₁₀ concentrations over the MASP. We found that the aerosol model used by MAIAC in the AOD retrievals does not adequately represent the aerosol properties of that region. However, the proposed prediction models that did not consider AOD data were able to explain, respectively, about 60% and 70% of the variability of diurnal and daily PM_{2.5} concentrations over the MASP. The performance of the prediction models for daily PM_{2.5} concentrations were quite satisfactory and comparable to the performance of the prediction models developed in Switzerland, Israel and Mexico City, which used AOD data from MAIAC. In addition, the prediction models that considered the two correlation sources had about 5% greater predictive power than the models that considered only the spatial correlation, and also generated more accurate estimates. The maps with the annual of the estimates of diurnal and daily PM_{2.5} concentrations over the MASP were made available in an open access repository. Those estimates can be used as an approximation for the

exposure to air pollution of the inhabitants in the MASP, contributing to epidemiological studies that aim to understand their impacts on the population's health.

Descriptors: Air pollution; Particulate matter; PM_{2.5}; PM₁₀; Atmospheric Aerosol; Aerosol Optical Depth; Air quality models; Remote sensing technology; Satellite imagery; Prediction models; Statistical models; Linear models; Environmental exposure.

Sumário

1	Introdução	1
1.1	Poluição atmosférica e material particulado	3
1.2	Efeitos da poluição atmosférica	4
1.2.1	Efeitos na saúde	4
1.2.2	Efeitos no ambiente e no clima	6
1.2.3	Efeitos econômicos	7
1.3	Avaliação da exposição humana à poluição atmosférica	8
1.4	Legislação e padrão da qualidade do ar no Brasil	12
1.5	Monitoramento da qualidade do ar no Brasil	15
1.6	Monitoramento dos aerossóis via sensoriamento remoto	18
1.7	Modelo Linear Misto (MLM) e a <i>Metodologia de Harvard</i> para a predição da concentração de $MP_{2.5}$ na superfície	20
2	Objetivos	23
2.1	Objetivo Geral	23
2.2	Objetivos Específicos	23
3	Métodos	25
3.1	Período e região do estudo	25
3.2	Dados primários e seu processamento	26
3.2.1	Concentrações de material particulado	27
3.2.2	Preditores espaço-temporais do $MP_{2.5}$	30
3.2.2.1	Dados de sensoriamento remoto	30
3.2.2.2	Dados meteorológicos	32
3.2.2.2.1	Altura da Camada Limite Planetária	34
3.2.3	Preditores espaciais do $MP_{2.5}$	37
3.2.3.1	Altitude	37
3.2.3.2	Uso e ocupação do solo	38
3.2.3.3	Emissões veiculares e densidade viária	40
3.3	Análise dos dados	45
3.3.1	Primeiro estágio da <i>Metodologia de Harvard</i> para a predição da concentração de $MP_{2.5}$	45
3.3.2	Relação entre as concentrações de $MP_{2.5}$ e MP_{10} medidos na superfície e a POA do MODIS-MAIAC na RMSP	46
3.3.2.1	Seleção das estações de monitoramento	47
3.3.2.2	Avaliação dos efeitos da ACLP, da UR e da sazonalidade na relação POA-MP	48
3.3.2.3	Avaliação do modelo de aerossol utilizado pelo algoritmo MAIAC na aquisição de dados de POA sobre a RMSP	50
3.3.3	Modelo proposto para a predição das concentrações de $MP_{2.5}$ na RMSP	51
3.3.3.1	Primeiro estágio	52
3.3.3.1.1	Seleção de variáveis nos modelos ajustados no primeiro estágio	55

3.3.3.1.2	Comparação dos modelos ajustados no primeiro estágio	55
3.3.3.1.3	Validação e avaliação do desempenho dos modelos ajustados no primeiro estágio	56
3.3.3.2	Segundo estágio	57
3.4	Questões Éticas	57
4	Resultados	59
4.1	Primeiro estágio da <i>Metodologia de Harvard</i> para a predição da concentração de $MP_{2.5}$ na RMSP	59
4.2	Relação entre as concentrações de $MP_{2.5}$ e MP_{10} medidas na superfície e a POA obtida do MODIS-MAIAC na RMSP	61
4.2.1	Correlação entre as concentrações de $MP_{2.5}$ e MP_{10} na superfície e os valores de POA do MODIS-MAIAC na RMSP	61
4.2.2	Avaliação dos efeitos da ACLP, da UR e da sazonalidade na relação POA-MP	62
4.2.3	Avaliação do modelo de aerossol utilizado pelo algoritmo MAIAC na aquisição de dados de POA sobre a RMSP	65
4.3	Modelo proposto para a predição das concentrações de $MP_{2.5}$ na RMSP	68
4.3.1	Primeiro estágio	68
4.3.2	Segundo estágio	77
5	Discussão	83
6	Conclusões	99
Apêndice A	Estudo da razão $MP_{2.5}/MP_{10}$ na RMSP no período 2000-2014	103
Apêndice B	<i>Script</i> para o ajuste dos modelos apresentados na seção 1.7	107
Apêndice C	Equações e interpretação dos parâmetros dos modelos	109
Apêndice D	<i>Script</i> para o ajuste do modelo que considera apenas a correlação espacial	111
Apêndice E	<i>Script</i> para o ajuste do modelo que considera as duas fontes de correlação	121
Apêndice F	Predição das concentrações de $MP_{2.5}$ e variância do erro de predição	127
F.1	Variância do erro de predição no Modelo de Harvard	131
F.1.1	Variância do erro de predição marginal	131
F.1.2	Variância do erro de predição condicional	132
F.2	Variância do erro de predição no modelo proposto	133
F.2.1	Variância do erro de predição marginal	133
F.2.2	Variância do erro de predição condicional	134
Apêndice G	Concentrações diárias observadas e preditas pelo modelo E1A1	135
Apêndice H	Concentrações diárias observadas e preditas pelo modelo E1A2	149
Apêndice I	Concentrações diurnas observadas e preditas pelo modelo E1A1	163
Apêndice J	Concentrações diurnas observadas e preditas pelo modelo E1A2	177
Apêndice K	Mapas das concentrações de $MP_{2.5}$ para a RMSP representados na mesma escala	191

Apêndice L Mapas das concentrações de MP _{2.5} para a Cidade de São Paulo	195
Apêndice M Mapas das concentrações de MP _{2.5} para a Cidade de São Paulo representados na mesma escala	199
Anexo 1 Artigo publicado na Revista <i>Atmospheric Environment</i>	205
Anexo 2 Parecer do Comitê de Ética e Pesquisa	217
Referências	219

1. Introdução

Desde a revolução industrial, enormes contingentes populacionais vêm sendo severamente atingidos pelos impactos ambientais (poluição sonora, do solo, da água e do ar) causados pelo rápido e descontrolado processo de crescimento econômico (Saldiva, 2018).

A partir da metade do século XX, a sociedade percebeu a gravidade da situação e iniciou um processo de discussão de políticas ambientais que pudessem compatibilizar esse crescimento com a qualidade de vida e manutenção da saúde das populações. No bojo dessa discussão, a Organização Mundial da Saúde (OMS) vem realizando esforços na compilação do conhecimento científico produzido principalmente na área acadêmica, para a elaboração de recomendações que guiem o desenvolvimento das políticas públicas ambientais (Saldiva, 2018).

No caso da poluição do ar, além da identificação dos poluentes mais relevantes em termos de impacto na saúde humana, também têm sido elaboradas recomendações de padrões ambientais que garantam o menor impacto à saúde. Na atualidade essas recomendações sugerem focar a atenção em dois poluentes atmosféricos: o material particulado (MP), principalmente na fração fina, e o ozônio (Saldiva, 2018). No presente estudo o foco será na fração fina do material particulado, também conhecida como MP_{2.5}.

Conhecer o padrão espaço-temporal da concentração dos poluentes atmosféricos é essencial para a elaboração de políticas de saúde pública e meio ambiente, além de ser um importante fator para prever alguns desfechos de saúde dos habitantes de uma área urbana. A principal forma de estimar a exposição à poluição atmosférica é por meio dos dados fornecidos por estações de monitoramento da qualidade do ar. Porém, os custos elevados de aquisição e principalmente de manutenção das redes de monitoramento da qualidade do ar tornam esparsa a distribuição espacial de suas estações. Nos países em desenvolvimento, a situação do monitoramento da qualidade do ar é bastante precária (Fajersztajn et al., 2014; WHO, 2016) o que dificulta a caracterização da variação espaço-temporal dos poluentes atmosféricos. No Brasil, por exemplo, apenas os estados de São Paulo e Rio de Janeiro contam com redes de monitoramento da qualidade do ar comparáveis às dos Estados Unidos e Europa (Alves et al., 2014).

Além disso, por meio dos dados obtidos dessas estações, só é possível estimar a exposição à poluição atmosférica com maior precisão apenas para os habitantes que vivem ou trabalham próximos dessas estações. Dessa forma, se faz necessário o desenvolvimento de metodologias que tenham como objetivo estimar as concentrações de poluentes em regiões sem estações de monitoramento, possibilitando estimar a exposição aos poluentes para toda a população de uma determinada região.

Nesse sentido, o monitoramento da qualidade do ar via sensoriamento remoto por satélites tem auxiliado na caracterização espaço-temporal dos poluentes atmosféricos em escala

regional e global (Chu et al., 2003), devido à sua ampla cobertura espacial associada à realização de medidas diárias confiáveis (Wang and Christopher, 2003). Desde 2003, a integração dos dados obtidos das estações na superfície com os dados obtidos do monitoramento da qualidade do ar via satélites tem auxiliado os profissionais que trabalham em agências de gerenciamento da qualidade do ar nos Estados Unidos (Al-Saadi et al., 2005; Hoff and Christopher, 2009).

Uma das medidas obtidas do monitoramento da qualidade do ar via sensoriamento remoto por satélites é a AOD, do termo em inglês *Aerosol Optical Depth*, e que em português vamos nos referir como Profundidade Óptica do Aerosol (POA). Trata-se de uma medida que indica o quanto a radiação solar foi atenuada ao atravessar a coluna atmosférica devido às interações (absorção ou espalhamento) com os constituintes da atmosfera (gases e partículas) e com a superfície (Wallace and Hobbs, 2006; Brasseur and Prinn, 2003). As pesquisas relacionadas à aquisição de dados de POA via satélite começaram em meados de 1970 e Wang and Christopher (2003) foram os primeiros a avaliar os dados de POA do instrumento MODIS (*Moderate Resolution Imaging Spectroradiometer*) como preditores das concentrações de $MP_{2.5}$ na superfície. Desde então, estudos desse tipo foram realizados em diferentes regiões do mundo, e a POA, no geral, mostra-se como uma importante preditora das concentrações de $MP_{2.5}$ na superfície (Chu et al., 2016).

No Brasil, poucos são os estudos que avaliaram os dados de POA como preditores das concentrações de $MP_{2.5}$ ou MP_{10} (Natali, 2008) bem como os que desenvolveram metodologias visando prever essas concentrações na superfície (Habermann and Gouveia, 2012). No entanto, muitos são os estudos que demonstram os efeitos deletérios das concentrações de poluentes na saúde dos habitantes, sendo predominantes os realizados na Cidade de São Paulo (Braga et al., 2001; Ribeiro and Cardoso, 2003; Saldiva et al., 1994a,b; Veras et al., 2010; Ribeiro et al., 2019; Takano et al., 2019).

Esta tese foi vinculada ao projeto temático FAPESP - Uso de modernas técnicas de autópsia na investigação de doenças humanas (MODAU, Processo FAPESP 2013/21728-2), mais especificamente ao subprojeto *METROHEALTH (Urban Health and Autopsy based approach)*, que teve como objetivo principal relacionar os desfechos de saúde observados nas autópsias realizadas nos indivíduos incluídos no MODAU com os efeitos da exposição à poluição atmosférica e com os hábitos de vida desses indivíduos.

Para isso, um dos objetivos do projeto *METROHEALTH*, e que foi o objetivo principal da tese, foi a calibração e o aprimoramento de um modelo estatístico para a predição das concentrações ambientais diárias de $MP_{2.5}$ na RMSP. Dessa forma, as estimativas de $MP_{2.5}$ obtidas do modelo podem ser utilizadas como uma aproximação para a exposição à poluição atmosférica dos indivíduos autopsiados no Projeto MODAU ou em estudos epidemiológicos futuros.

1.1 Poluição atmosférica e material particulado

A poluição atmosférica ocorre quando espécies estão presentes no ar em concentrações suficientemente altas para produzir efeitos agudos ou crônicos na saúde humana, além de causar impactos negativos para os animais, para a vegetação ou para os materiais (Seinfeld and Pandis, 2006; Wallace and Hobbs, 2006; Tan, 2014). Em princípio, tais espécies podem estar em qualquer fase: partículas de poeira sólida, gotículas de líquido ou na fase gasosa em suspensão no ar (Tan, 2014).

Essas espécies têm origens naturais ou antropogênicas, podendo ser emitidas diretamente por suas fontes (espécies primárias) ou serem produzidas por reações químicas de duas ou mais espécies primárias ou ainda por reações dessas espécies com outros constituintes da atmosfera (espécies secundárias) (Tan, 2014). Como exemplo de espécies primárias podemos citar o dióxido de enxofre (SO_2), os óxidos nítricos (NO_X), os compostos orgânicos voláteis (COVs) e o monóxido de carbono (CO). O ozônio (O_3) e o ácido sulfúrico (H_2SO_4) são exemplos de espécies secundárias (Tan, 2014). E existem aquelas que podem ser tanto emitidas por suas fontes quanto formadas na atmosfera, como o MP e os aldeídos.

O MP é a componente particulada do aerossol atmosférico, definido como uma mistura de partículas finas (sólidas ou líquidas), que por serem muito pequenas ($<100 \mu m$) ficam suspensas em um gás. No entanto, o termo aerossol atmosférico tem sido utilizado como sinônimo de MP, o que também faremos ao longo de todo o texto (Seinfeld and Pandis, 2006; Finlayson-Pitts and Pitts Jr, 1999).

O MP origina-se tanto de processos naturais, como poeira transportada pelo vento, erupções vulcânicas ou pólen, quanto de processos antrópicos, como a combustão de combustíveis fósseis, emissões industriais ou a queima de biomassa (geralmente relacionada ao manejo agropecuário). Os aerossóis gerados pelos processos citados anteriormente são exemplos de aerossol primário. Como exemplos de aerossol secundário podemos citar o sulfato, o nitrato, o amônio e os orgânicos secundários, resultantes principalmente de processos do tipo conversão gás-partícula, resultado de reações físico-químicas que ocorrem na atmosfera (Seinfeld and Pandis, 2006).

Uma característica importante do MP é o seu diâmetro (D_p), pois este está relacionado ao seu tempo de permanência na atmosfera. De acordo com o D_p , o MP pode ser classificado como partículas totais em suspensão (PTS, $D_p < 50 \mu m$), partículas inaláveis (MP_{10} , $D_p < 10 \mu m$), partículas inaláveis finas ($MP_{2.5}$, $D_p < 2.5 \mu m$) e fumaça (associada ao MP na atmosfera proveniente dos processos de combustão e está diretamente relacionada ao teor de fuligem na atmosfera) (CETESB, 2019a). Quanto menores as partículas, maior o tempo de permanência na atmosfera. E em concentrações que excedam as naturais, além de alterarem o balanço radiativo atmosférico, potencializam a sua penetração no sistema respiratório, trazendo prejuízos à saúde (World Health Organization, 2006).

1.2 Efeitos da poluição atmosférica

1.2.1 Efeitos na saúde

Três eventos atípicos de poluição atmosférica ocorridos na primeira metade do século XX, conhecidos como “os grandes desastres da poluição do ar”, mudaram as atitudes da sociedade frente à problemática da poluição do ar. Os locais e as datas desses eventos foram: vale do Rio Meuse, Bélgica (1930), vale do Rio Monogahela, Pennsylvania (1948) e vale do Rio Tamisa, Londres (1952) (Phalen and Phalen, 2011).

Nos três eventos, as emissões de poluentes (emissões de queima de carvão utilizado no aquecimento das residências combinadas com as emissões industriais e de tráfego veicular/-trens) em conjunto com a topografia da região (vales de rios) e as condições meteorológicas desfavoráveis para a dispersão de poluentes (baixas temperaturas, pouco vento e a presença de nevoeiro intenso), resultaram em 63, entre 18 e 20 e entre 3000 e 4000 mortes em excesso, na Bélgica, na Pennsylvania e em Londres, respectivamente (Phalen and Phalen, 2011).

O evento de Londres tornou-se o mais conhecido, pois foi a primeira vez que ocorreu em uma cidade grande e desenvolvida para os padrões da época. Após o ocorrido, o Parlamento Britânico promulgou em 1956 o *British Clean Air Act*, que limitou o uso de carvão para aquecimento doméstico. O *British Clean Air Act* desencadeou outras iniciativas no Reino Unido e no mundo, que resultaram em ações regulatórias, dentre elas o *U.S. National Clean Air Act*, em 1963 (Phalen and Phalen, 2011).

Logo após o evento ocorrido em Londres, começaram os ensaios *in vitro* e os experimentos em animais e humanos, que estudaram os efeitos da exposição a diferentes concentrações de poluentes. Iniciaram-se também diversos estudos epidemiológicos, cujas conclusões foram criticadas, pois falharam ao não considerar variáveis de controle, como o tabagismo (Phalen and Phalen, 2011).

Em 1993, foi publicado o primeiro estudo prospectivo que teve como objetivo avaliar os efeitos da poluição do ar na saúde (Dockery et al., 1993). Esse estudo, conhecido como *Harvard Six Cities* (Ferris et al., 1979), conduzido por pesquisadores da Universidade de Harvard, teve início em 1974 e acompanhou 8111 indivíduos de seis cidades dos Estados Unidos por, em média, 16 anos. Controlando fatores como gênero, idade, índice de massa corpórea, nível de escolaridade e tabagismo, Dockery et al. (1993) encontraram uma associação estatisticamente significativa entre poluição do ar (principalmente partículas finas) e mortalidade.

Desde o final da década de 1990, houve um crescimento no número de pesquisas, principalmente nos Estados Unidos e na Europa, que trouxeram evidências dos efeitos da exposição ao MP e a outros poluentes na saúde. Em 2005, a OMS produziu um relatório que atualizou os valores das diretrizes para os quatro poluentes atmosféricos mais comuns - MP, O_3 , NO_2 e SO_2 , além de apresentar diversos estudos que relacionam esses poluentes e diferentes desfechos de morbimortalidade (World Health Organization, 2006).

As concentrações de $MP_{2.5}$ são comumente utilizadas como uma aproximação (*proxy*) para a exposição à poluição do ar. Mesmo em concentrações muito baixas, o $MP_{2.5}$ causa impactos na saúde, de forma que alguns estudos sugerem que possa não existir um limite mínimo abaixo do qual tais impactos não existam. Em 2014, 92% da população mundial estava exposta às concentrações de $MP_{2.5}$ acima do nível médio anual de $10 \mu g/m^3$ recomendado pela OMS (WHO, 2016).

No Brasil, o sexto país que mais publica artigos relacionados à poluição atmosférica (Vormittag et al., 2013a), os estudos relacionando as concentrações de poluentes aos efeitos à saúde começaram na década de 1990 e concentram-se no estado de São Paulo, sendo pioneiros os estudos de Saldiva et al. (1992, 1994a,b).

Saldiva et al. (1992), estudaram os efeitos da exposição aos poluentes atmosféricos em ratos, na Cidade de São Paulo. Após seis meses de exposição, os animais apresentaram lesões no sistema respiratório, como hiperplasia de células secretoras e muco mais rígido, alterações que comprometem o mecanismo de autolimpeza da mucosa nasal.

Controlando fatores como a temperatura e a umidade relativa do ar, Saldiva et al. (1994a) encontraram uma associação significativa entre os níveis de NO_X e a mortalidade por doenças respiratórias em crianças (< 5 anos). A probabilidade de óbito por causas respiratórias em crianças, considerando a concentração a média de NO_X igual a 0,127 ppm em São Paulo durante o período do estudo, foi estimada em cerca de 30%.

Saldiva et al. (1994b) encontraram uma associação entre a mortalidade (todas as causas) e as concentrações de MP_{10} , SO_2 , NO_X e CO em idosos (> 65 anos), sendo que a associação com o MP_{10} foi a mais significativa. Um aumento de $10 \mu g/m^3$ na concentração de MP_{10} aumentou a mortalidade em 1,3%.

Depois dos estudos de Saldiva et al. (1992, 1994a,b), outros trabalhos mostraram a associação entre as concentrações de poluentes (MP_{10} , O_3 , SO_2 , NO_2 e CO) e diferentes desfechos de morbidade (Gouveia and Fletcher, 2000; Nascimento et al., 2006; Amâncio and Nascimento, 2012; Tuan et al., 2015; Gouveia et al., 2017; Nascimento et al., 2017) e mortalidade (Conceição et al., 2001) relacionados ao sistema respiratório, principalmente em crianças, na Cidade de São Paulo. Dentre os artigos citados anteriormente, o MP_{10} não se mostrou associado ao número de hospitalizações diárias por pneumonia apenas no estudo de Tuan et al. (2015).

Outros trabalhos conduzidos na Cidade de São Paulo também encontraram associações entre as concentrações de poluentes, especialmente o MP_{10} , e desfechos não relacionados ao sistema respiratório: baixo peso ao nascer (Gouveia et al., 2004; Medeiros and Gouveia, 2005; Romão et al., 2013), mortalidade neonatal (o óbito ocorre até os 28 dias após o nascimento) (Lin et al., 2004), mortalidade por causas cardiovasculares em adultos (Pereira Filho et al., 2008; Gouveia et al., 2017) e incidência de câncer por causas respiratórias (Ribeiro et al., 2019).

Fora da Cidade de São Paulo, predominam os trabalhos realizados no interior do estado e na região da Amazônia brasileira. Eles estudaram os efeitos das emissões de $MP_{2.5}$ e MP_{10} que

ocorrem durante o período de queima de biomassa, na morbidade por causas respiratórias. No interior do estado de São Paulo, onde as queimadas ocorrem principalmente entre os meses de maio e outubro durante a colheita de cana de açúcar, podemos citar os trabalhos de Arbex et al. (2007), Cançado et al. (2006), Silva et al. (2010) e Pestana et al. (2017). Já na região da Amazônia brasileira, onde as queimadas ocorrem entre agosto e novembro (estação seca), devido principalmente à alteração no uso e ocupação do solo, podemos citar os trabalhos de Carmo et al. (2010), Ignotti et al. (2010), Gonçalves et al. (2012), Andrade et al. (2012a) e Andrade Filho et al. (2017).

Os trabalhos de Fernandes et al. (2020) e Andreão and Albuquerque (2020) apresentam respectivamente revisões detalhadas dos estudos epidemiológicos realizados no Brasil que relacionaram as concentrações de MP_{10} e $MP_{2.5}$ com desfechos de saúde.

A revisão realizada por Andreão and Albuquerque (2020) incluiu 16 estudos que relacionaram as concentrações de $MP_{2.5}$ com doenças dos sistemas respiratório (13), circulatório (1) e cardiovascular (2). A metade dos estudos (8) foi realizada no estado de São Paulo, em cidades fora da RMSP.

1.2.2 Efeitos no ambiente e no clima

Além dos efeitos na saúde humana, as emissões antropogênicas causam redução da visibilidade nas áreas urbanas, chuva ácida e perturbam o balanço de radiação da Terra (Seinfeld and Pandis, 2006).

Quando as taxas de emissão ou formação de poluentes excedem as taxas que os poluentes são dispersos ou removidos da atmosfera, observa-se graves episódios de poluição do ar, sendo a redução da visibilidade uma percepção direta da má qualidade do ar (Wallace and Hobbs, 2006).

Os gases e partículas presentes em excesso na atmosfera podem absorver, mas principalmente dispersar a luz, o que leva à diminuição da visibilidade. As partículas menores ($0,1 - 1,1 \mu m$) são as mais eficientes em dispersar a luz (Seinfeld and Pandis, 2006). A composição das partículas também é uma fator importante para a redução da visibilidade. As partículas finas de sulfato e nitrato podem permanecer suspensas no ar por um longo período de tempo, dando à poluição uma cor marrom-amarelada (Tan, 2014).

A precipitação é a principal forma de remoção dos poluentes da atmosfera, e durante esse processo pode ocorrer a chuva ácida (Wallace and Hobbs, 2006). Uma gota de chuva em uma atmosfera livre de poluentes tem pH^1 de 5,6 (Seinfeld and Pandis, 2006). No ar poluído, devido à incorporação de poluentes gasosos e particulados na chuva, o pH da água da chuva pode ser muito menor que 5,6, dando origem à chuva ácida (Wallace and Hobbs, 2006).

¹O pH de um líquido é definido como $pH = -\log[H_3O^+(aq)]$ em que $H_3O^+(aq)$ é a concentração (em mol por litro) dos íons de H_3O^+ no líquido. Observe que a mudança de uma unidade no pH corresponde a uma mudança de um fator de 10 no $H_3O^+(aq)$. Soluções ácidas têm $pH < 7$ e soluções básicas têm $pH > 7$ (Wallace and Hobbs, 2006).

O gases mais importantes que levam à acidificação da chuva são o SO_2 e o NO_X . Eles são convertidos em ácido sulfúrico e ácidos nítrico e nitroso, respectivamente, além de as emissões de NO_X contribuírem para a formação do O_3 (Seinfeld and Pandis, 2006; Tan, 2014).

Quando chega na superfície, a chuva ácida afeta o solo, as florestas e os corpos de água doce. A chuva ácida altera a biologia dos micróbios que vivem no solo, remove os minerais e os nutrientes necessários para a manutenção do ecossistema, o que pode alterar o crescimento de plantas e florestas (Tan, 2014). Ao cair nos corpos de água, a chuva ácida promove a acidificação das águas superficiais, trazendo danos aos ecossistemas aquáticos (Seinfeld and Pandis, 2006).

A chuva ácida também pode afetar a aparência de prédios e monumentos históricos ao corroer o seus materiais. Um exemplo disso, é a perda de brilho das esculturas, devido à reação química da chuva ácida com os compostos de cálcio presentes nas esculturas (Tan, 2014).

As partículas do aerossol atmosférico desempenham um papel importante na regulação do clima e da qualidade do ar, afetando diretamente e indiretamente o balanço radiativo atmosférico da Terra. Dependendo de seu tamanho e composição, essas partículas podem espalhar ou absorver a radiação solar, alterando a temperatura da superfície (efeito direto) ou alterando o tempo de vida e o albedo das nuvens, diminuindo a precipitação (efeito indireto) (Stocker et al., 2013).

Além disso, partículas menores têm maior potencial de serem transportadas por longas distâncias, afetando assim a química e a física da atmosfera, em escala regional e até global. Como exemplos, podemos citar o transporte da fumaça resultante das queimadas que ocorrem na floresta Amazônica e região Central para diversas regiões do Brasil (Mariano et al., 2010; de Miranda et al., 2017; Yamasoe et al., 2017; Vara-Vela et al., 2018), o transporte da poluição originária na região conhecida como Círculo de Fogo, no Oceano Pacífico, para a América do Norte (Wilkening et al., 2000; Bailey et al., 2000; Hadley et al., 2007) e o transporte de poeira do Deserto do Saara para a região do Caribe e norte da América do Sul (Prospero, 1999; Yu et al., 2015; Chouza et al., 2016; van der Does et al., 2020).

1.2.3 Efeitos econômicos

Os efeitos econômicos de uma doença estão muito relacionados aos seus impactos na saúde da população, sejam eles devido aos óbitos decorrentes ou ao grau de adoecimento dos indivíduos, esse último impactando diretamente a qualidade de vida. Óbitos e adoecimentos implicam menos produção, consumo e geração de renda, além dos custos com o tratamento da doença em si (WHO et al., 2009).

No caso da poluição atmosférica, de acordo com a Organização para Economia, Cooperação e Desenvolvimento (OECD), em 2050, a exposição ao MP será a principal causa de mortes prematuras nos países membros, alcançando cerca de 3,6 milhões de mortes anuais

(OECD, 2012).

Em 2012, a poluição por MP foi responsável por cerca de 3 milhões de mortes no mundo, sendo que aproximadamente 87% docorreram em países de média e baixa renda, o que representa 82% da população mundial. Na América Latina, o Brasil, com 26.241 mortes, foi o país onde mais ocorreram mortes atribuídas às concentrações de MP (WHO, 2016).

De acordo com o estudo conduzido pelo Instituto Saúde e Sustentabilidade, em 2011, na Cidade de São Paulo, a poluição atmosférica matou 4.655 indivíduos o que corresponde a seis vezes o número de mortes por câncer de próstata ou AIDS, três vezes e meia o número de mortes por câncer de mama e três vezes o número de mortes por acidentes de trânsito (Vormittag et al., 2013b).

Traduzindo em número de anos perdidos por morte precoce e vividos com incapacidade (do termo em inglês *Disability-Adjusted Life Year* (DALY)) atribuíveis à poluição atmosférica, em 2011, na cidade, na região metropolitana e no estado de São Paulo, foram respectivamente 40.811, 76.497 e 159.422 anos, valorados em mais de 302 bilhões, 566 bilhões e 1 trilhão de dólares na cidade, na região metropolitana e no estado de São Paulo, respectivamente (Vormittag et al., 2013b).

Também existem os benefícios econômicos decorrentes do investimento em sistemas de saúde e políticas ambientais (WHO et al., 2009). A implantação de políticas para regulação e controle da qualidade do ar decorrentes do *Clean Air Act* nos Estados Unidos, implicaram gastos acima de 520 bilhões de dólares entre 1970 e 1990, que trouxeram benefícios estimados acima dos 21 trilhões de dólares (Christopherson, 2012).

Extensões do estudo *Harvard Six Cities* mostraram os impactos positivos das políticas de controle da qualidade do ar implantadas pela *United States Environmental Protection Agency* (USEPA): a diminuição das concentrações de MP_{2.5} foi acompanhada pela diminuição da mortalidade geral (todas as causas) (Laden et al., 2006; Lepeule et al., 2012).

1.3 Avaliação da exposição humana à poluição atmosférica

Como as concentrações dos poluentes e as atividades humanas variam no tempo e no espaço, avaliar a exposição de um indivíduo a um poluente atmosférico é complexo, já que é influenciada pela interação do indivíduo com os ambientes internos e externos (microambientes) que frequenta (Zou et al., 2009; Millar et al., 2010a). A Figura 1.1, adaptada de Zou et al. (2009), esquematiza o processo de exposição à poluição atmosférica. A exposição aos poluentes em ambientes internos é a que tem maior potencial de impacto à saúde uma vez que a maioria dos indivíduos passa a maior parte do seu tempo nesses ambientes. No entanto, é muito comum utilizar apenas as estimativas das concentrações ambientais como uma aproximação para a exposição individual (WHO, 2016).

Figura 1.1: Ilustração do processo de exposição à poluição atmosférica. Baseado e adaptado dos esquemas apresentados em Zou et al. (2009) e em EEA (2016).

No caso dos ambientes externos, uma vez emitidos, os poluentes podem ser transportados, reagirem entre si ou com outros constituintes da atmosfera (por exemplo, formação dos aerossóis secundários ou do ozônio), ou mesmo sofrerem deposição (por exemplo, deposição úmida pela chuva). No caso dos ambientes internos, ocorre a mesma dinâmica externa de “emissão, transformação e dispersão”, e soma-se a ela os poluentes externos que podem adentrar os ambientes internos através de portas e janelas, rachaduras ou mesmo infiltrando-se pela construção (Millar et al., 2010a).

Os trabalhos de Monn (2001), Zou et al. (2009), Millar et al. (2010a) e Millar et al. (2010b), entre outros, apresentam uma revisão dos métodos de avaliação da exposição humana à poluição atmosférica. Baseando-se nesses trabalhos, construímos a Figura 1.2, que esquematiza os métodos de avaliação da exposição individual.

Figura 1.2: *Métodos de avaliação da exposição.*

Os métodos diretos de avaliação da exposição são os que melhor refletem os níveis de exposição dos indivíduos, pois são capazes de mensurar a concentração do poluente no ambiente imediatamente antes da inalação (monitoramento pessoal via amostrador portátil) ou a resposta interna da exposição à alguma fonte específica (marcadores biológicos que podem ser detectados na urina, no sangue, cabelo, unhas, esfregaço nasal, etc). No entanto, esses métodos possuem a desvantagem de serem invasivos e caros (Millar et al., 2010a).

Os métodos indiretos, apesar de menos precisos, são formas relativamente eficientes e mais econômicas de se avaliar a exposição de indivíduos (Millar et al., 2010b).

A principal forma indireta de estimar a exposição à poluição atmosférica é por meio de dados do monitoramento das concentrações ambientais dos poluentes, fornecidos por estações em redes de monitoramento da qualidade do ar, que geralmente são mantidas pelos governos para fins de regulamentação (Millar et al., 2010b). Outra forma, é através do monitoramento das concentrações dos principais microambientes (internos e externos) que os indivíduos frequentam: dentro dos diferentes cômodos da residência, dentro dos veículos de locomoção (individual ou coletivo), no local de trabalho, em áreas externas ao local que vivem, trabalham ou praticam atividade física. Em conjunto com dados de atividades diárias dos indivíduos, geralmente coletados por meio de questionários/diários, o monitoramento de microambientes é essencial para avaliar a exposição de longo prazo, principalmente nos estudos retrospectivos (Monn, 2001).

Existe ainda a avaliação da exposição à poluição atmosférica por meio de modelos, que de uma forma geral, podem ser classificados em modelos estatísticos, modelos físico-químicos e modelos híbridos (modelos que combinam diferentes abordagens para estimar a exposição individual) (Monn, 2001; Zou et al., 2009; Millar et al., 2010b). O desenvolvimento desses modelos tem sido impulsionado pelos avanços tecnológicos que ocorreram nas últimas décadas, que possibilitaram o aumento da capacidade de processamento dos computadores, que

por sua vez aumentou a capacidade de processamento de dados (Brasseur and Prinn, 2003; Millar et al., 2010b).

Os modelos estatísticos tentam estabelecer relações entre dados mensurados (por exemplo, dados de estações de monitoramento) e variáveis explicativas (distância de fontes de emissão de poluentes, dados meteorológicos, topográficos e de uso e ocupação do solo) (Monn, 2001; Millar et al., 2010b). Como exemplo, podemos citar os modelos de proximidade, os diferentes métodos de interpolação espacial/geoestatísticos como o IDW (*Inverse Distance Weighting*) e o *Kriging*, e os modelos de regressão baseada em variáveis de uso e ocupação do solo (*land use regression*). Diferentes estudos que aplicaram esses modelos podem ser encontrados nas revisões apresentadas por Zou et al. (2009) e Millar et al. (2010b), por exemplo.

Os modelos físicos-químicos, mais conhecidos como Modelos de Transporte Químico (MTQs, do termo em inglês, *Chemical Transport Models* (CTMs)), tentam representar por meio de equações os principais processos físicos (transporte, dispersão e deposição) e químicos (por exemplo, formação de ozônio e de aerossol secundário) que ocorrem na atmosfera, que controlam as distribuições espacial e temporal de gases traços e partículas de aerossóis (Brasseur and Prinn, 2003).

No Brasil, de acordo com a revisão conduzida por Andreão and Albuquerque (2020), 13 dos 16 estudos epidemiológicos que relacionaram as concentrações de $MP_{2.5}$ com desfechos de saúde utilizaram o *Chemical Coupled Aerosol and Tracer Transport model to the Brazilian developments on the Regional Atmospheric Modeling System* (CCATT-BRAMS), desenvolvido pelo Centro de Previsão de Tempo e Estudos Climáticos/Instituto Nacional de Pesquisas Espaciais (CPTEC/INPE), para quantificar as concentrações diárias de $MP_{2.5}$. Já os modelos *WRF-Chem* (*Weather Research and Forecasting with Chemistry*) e o *Community Multi-Scale Air Quality* (CMAQ) foram utilizados para quantificar as concentrações de $MP_{2.5}$ nos estudos conduzidos por Andrade et al. (2015), Vara-Vela et al. (2018), Albuquerque et al. (2018) e Albuquerque et al. (2019), na Cidade de São Paulo.

Embora por meio dos modelos físico-químicos seja possível obter estimativas das concentrações dos poluentes com melhores resoluções espacial e temporal, suas principais desvantagens são a necessidade de dados das emissões (veiculares, industriais e naturais), que muitas vezes não estão disponíveis ou não estão suficientemente detalhados (Monn, 2001; Millar et al., 2010b), além de serem custosos do ponto de vista computacional.

No que diz respeito ao desempenho dos modelos estatísticos e físico-químicos, Hoek et al. (2008), que realizaram uma revisão dos estudos que utilizaram modelos de regressão baseada em variáveis de uso e ocupação do solo para avaliar a variação espacial de poluentes, encontraram que os modelos de regressão baseada em variáveis de uso e ocupação do solo geralmente apresentam melhor desempenho do que os métodos geoestatísticos e apresentam desempenho semelhante ou superior ao desempenho dos MTQs para explicar a variação espacial dos poluentes. No entanto, como as variáveis de uso e ocupação do solo geralmente não variam no tempo, os modelos de regressão baseada em variáveis de uso e ocupação do

solo só podem ser empregados para avaliar os efeitos da exposição de longo prazo (efeitos crônicos) na saúde humana (Kloog et al., 2012, 2014; Lee et al., 2016).

Um avanço nesse sentido é a metodologia proposta por pesquisadores da Faculdade de Saúde Pública de Harvard, que além de permitir realizar predições das concentrações de MP em regiões sem estações de monitoramento, permite a avaliação dos efeitos da exposição para desfechos de curto e de longo prazo (Lee et al., 2011; Kloog et al., 2011, 2012; Chudnovsky et al., 2014b). As bases dessa metodologia são: utilizar o Modelo Linear Misto (MLM, Laird and Ware (1982)), um modelo estatístico que permite considerar a variação diária dos fatores que influenciam a variação das concentrações de MP (por exemplo, as condições meteorológicas); e utilizar a relação já conhecida entre as concentrações de MP_{2.5} obtidas de estações de monitoramento e a POA obtida de satélites, que será discutida na seção 1.6.

Os modelos que combinam as concentrações dos ambientes internos e externos para a avaliação da exposição individual, sejam elas mensuradas ou modeladas, são chamados de modelos de inalação e modelos híbridos por Zou et al. (2009) e Millar et al. (2010b), respectivamente. Por meio desses modelos, é possível estimar os níveis de exposição de forma mais precisa, pois eles exigem informações detalhadas em nível individual, como tempo de permanência em cada microambiente, peso, idade, etc. Como exemplos podemos citar o *Simulation of Human Activity and Pollutant Exposure* (SHAPE, Ott et al. (1988)), o *Stochastic Human Exposure and Dose Simulation* (SHEDS, Burke et al. (2001)) e o *Air Pollutants Exposure* (APEX, Zartarian et al. (2000)).

1.4 Legislação e padrão da qualidade do ar no Brasil

No Brasil, a Resolução número 5 do Conselho Nacional do Meio Ambiente (CONAMA), publicada em 15 de junho de 1989 (Resolução CONAMA nº 5/1989), instituiu o Programa Nacional de Controle da Qualidade do Ar (PRONAR),

“como um dos instrumentos básicos da gestão ambiental para proteção da saúde e bem estar das populações e melhoria da qualidade de vida com o objetivo de permitir o desenvolvimento econômico e social do país de forma ambientalmente segura, pela limitação dos níveis de emissão de poluentes por fontes de poluição atmosférica com vistas a: uma melhoria na qualidade do ar; o atendimento aos padrões estabelecidos; o não comprometimento da qualidade do ar em áreas consideradas não degradadas.”

Por meio de um dos instrumentos de gestão previstos no PRONAR, em 28 de junho de 1990, foi publicada a Resolução CONAMA nº 3/1990, que além de definir os padrões de qualidade do ar a serem seguidos em todo o território nacional, atribuiu aos estados a responsabilidade de implantarem as redes de monitoramento. Porém, em 19 de novembro de 2018, foi publicada a Resolução CONAMA nº 491/2018, que revoga a Resolução CONAMA

nº 3/1990, e estabelece os novos padrões de qualidade do ar, que agora têm como base as recomendações da OMS publicadas em 2005 ([World Health Organization, 2006](#)). Os atuais padrões de qualidade do ar estão em vigência desde 2013 no estado de São Paulo, por meio do Decreto Estadual nº 59.113/2013.

A Resolução CONAMA nº 491/2018 estabelece padrões de qualidade do ar intermediários (PI), que são os valores temporários a serem cumpridos de forma gradativa até que a qualidade do ar atinja o padrão de qualidade do ar final (PF), que são os valores preconizados pela OMS. Em comparação aos padrões estabelecidos na Resolução CONAMA nº 3/1990, os novos padrões estabelecidos pela Resolução CONAMA nº 491/2018 são mais protetivos (menores) para MP_{10} , SO_2 , NO_2 e fumaça. Para o O_3 , além de o padrão ter diminuído, o tempo de amostragem passou de 1h para 8h. Para o CO , não existem mais os padrões intermediários, mas apenas o padrão final igual a 9 ppm e também deixa de existir o tempo de amostragem de 1h, permanecendo apenas o de 8h. Para as partículas totais em suspensão, da mesma forma que para o CO , passa a existir apenas o padrão final, porém este é igual ao padrão primário da resolução anterior. Além disso, a nova resolução inclui o monitoramento do $MP_{2.5}$ e do chumbo (Pb). A comparação dos padrões de qualidade do ar estabelecidos pelas Resoluções CONAMA nº 3/1990 e CONAMA nº 491/2018 é apresentada na Tabela 1.1.

Tabela 1.1: *Padrões nacional de qualidade do ar estabelecidos pelas Resoluções CONAMA nº 3, de 28/06/1990 e nº 491, de 19/11/2018.*

Poluente	Resolução de 1990			Resolução de 2018				
	Tempo de Amostragem	Padrão Primário	Padrão Secundário	Tempo de Amostragem	PI 1	PI 2	PI 3	PF
Partículas inaláveis (MP_{10})	24 horas MAA ¹	150 50	150 50	24 horas MAA ¹	120 40	100 35	75 30	50 20
Partículas inaláveis finas ($MP_{2.5}$)	24 horas MAA ¹	- -	- -	24 horas MAA ¹	60 20	50 17	37 15	25 10
Dióxido de enxofre (SO_2)	24 horas MAA ¹	365 80	100 40	24 horas MAA ¹	125 40	50 30	30 20	20 -
Dióxido de nitrogênio (NO_2)	1 hora MAA ¹	320 100	190 100	1 hora ² MAA ¹	260 60	240 50	220 45	200 40
Ozônio (O_3)	1 hora	160	160	8 horas ³	140	130	120	100
Monóxido de carbono (CO)	1 hora	40000	40000	1 hora	-	-	-	-
		35 ppm	35 ppm		-	-	-	-
	8 horas	10000 9 ppm	10000 9 ppm	8 horas	-	-	-	9 ppm
Fumaça	24 horas	150	100	24 horas	120	100	75	50
	MAA ¹	60	40	MAA ¹	40	35	30	20
Partículas Totais em Suspensão (PTS)	24 horas	240	150	24 horas	-	-	-	240
	MGA ⁴	80	60	MGA ⁴	-	-	-	80
Chumbo (Pb) ⁵	MAA ¹	-	-	MAA ¹	-	-	-	0,5

Reproduzido de *CETESB (2018)* e *CETESB (2019a)*. Nota: os padrões vigentes estão em negrito; 1-Média aritmética anual; 2-Média horária; 3-Máxima média móvel obtida no dia; 4-Média geométrica anual; 5-Medido nas partículas totais em suspensão; PI-Padrão intermediário; PF-Padrão final.

Vale enfatizar que na Cidade de São Paulo e na RMSP, os padrões anuais para $MP_{2.5}$ ($10 \mu g/m^3$) e MP_{10} ($20 \mu g/m^3$) estabelecidos pela OMS, que são mais restritivos do que os atuais padrões nacionais, foram ultrapassados em todos os anos dos períodos 2000-2019 e 2005-2015, respectivamente, de acordo com os estudos de [Andreão and Albuquerque \(2020\)](#) e [Pacheco et al. \(2017\)](#). Os padrões diários para $MP_{2.5}$ ($25 \mu g/m^3$) e MP_{10} ($50 \mu g/m^3$) também são ultrapassados com frequência na RMSP, principalmente na estação seca (Maio-Outubro) ([Fernandes et al., 2020](#)).

O Instituto de Energia e Meio Ambiente (IEMA) faz o alerta de que “*sem criar melhores condições para gestão, adotar padrões de qualidade do ar mais restritivos, levará apenas ao aumento de registros de ultrapassagens*” ([IEMA, 2014](#)).

A ausência dos prazos para a alteração dos padrões intermediários definidos tanto pela Resolução CONAMA nº 491/2018 quanto pelo Decreto Estadual nº 59.113/2013 no estado de São Paulo, além de prolongar os efeitos deletérios causados pela poluição na saúde da população, acaba precocemente com vidas. Críticas referentes à ausência dos prazos para a alteração dos padrões intermediários são apresentadas nos trabalhos de [Andrade et al. \(2017\)](#), [Vormittag and Saldiva \(2017\)](#), [Fernandes et al. \(2020\)](#), [Andreão and Albuquerque \(2020\)](#), por exemplo.

1.5 Monitoramento da qualidade do ar no Brasil

O monitoramento é uma ferramenta fundamental para a avaliação da qualidade do ar ([WHO, 2000](#)), além de ser a base para a aplicação dos instrumentos de gestão ambiental ([IEMA, 2014](#)). Os dados obtidos do monitoramento fornecem informações sobre os níveis passado e presente da qualidade do ar, auxiliando o poder público no planejamento de políticas de saúde pública e meio ambiente ([Alves et al., 2014](#)). Também são a base para os estudos de dispersão da poluição atmosférica e para os estudos epidemiológicos que tentam entender os danos causados à saúde humana e aos sistemas ambientais devido à exposição ao ar poluído. Além disso, servem para avaliar se os padrões de qualidade do ar estabelecidos pela legislação em vigência estão sendo atendidos.

Apesar de a Resolução CONAMA nº 5/1989 ter previsto a constituição de uma rede básica de monitoramento da qualidade do ar, para a maioria dos estados brasileiros a legislação não foi cumprida. No ano de 2014 foram publicados dois estudos que compilaram informações sobre a configuração e a cobertura territorial das redes de monitoramento no Brasil: o “*1º Diagnóstico das redes de monitoramento da qualidade do ar no Brasil*” ([IEMA, 2014](#)), produzido pelos esforços do IEMA, dos órgãos estaduais de meio ambiente (OEMA) e do Ministério do Meio Ambiente (MMA), e o “*Monitoramento da Qualidade do Ar no Brasil*”, produzido pelo Instituto Saúde e Sustentabilidade ([Alves et al., 2014](#)). Ambos os estudos incluíram os estados cujas informações estavam disponíveis nas páginas eletrônicas dos órgãos ambientais de cada estado, independente da responsabilidade de gestão das

estações de monitoramento (pública e/ou privada). A principal diferença entre os estudos está no período em que as informações foram levantadas: 2012 (IEMA, 2014) e 2014 (Alves et al., 2014). Apesar de utilizarem metodologias um pouco diferentes, os achados em ambos os estudos são parecidos e nada animadores.

Com base nos dados fornecidos por esses estudos, construímos a Figura 1.3, que destaca as unidades federativas brasileiras que operavam redes de monitoramento, e a Tabela 1.2, que apresenta o número total de estações de monitoramento e especifica quantas delas monitoravam o MP_{10} e o $MP_{2.5}$ em cada estado. De acordo com os estudos conduzidos por IEMA (2014) e Alves et al. (2014), o número de unidades federativas que operavam redes de monitoramento da qualidade do ar no Brasil eram nove e onze, respectivamente. No entanto, nas conclusões do levantamento realizado por Alves et al. (2014), fez-se uma ressalva aos estados de Mato Grosso, Goiás e Sergipe, no sentido de que esses estados não deveriam ter sido incluídos no estudo devido ao número de estações e poluentes mensurados. Dessa forma, o número de unidades federativas que operavam redes de monitoramento encontrado pelos dois estudos é praticamente o mesmo. De acordo com os dois estudos, nenhum estado da região Norte do país possuía rede de monitoramento da qualidade do ar, enquanto que todos os estados da região Sudeste contavam com rede de monitoramento. Nos dois estudos, os estados de São Paulo e Rio de Janeiro foram os estados cujas redes possuíam o maior número de estações de monitoramento da qualidade do ar.

Figura 1.3: Unidades federativas no Brasil que operavam redes de monitoramento da qualidade do ar, de acordo com os estudos de (A) IEMA (2014) e (B) Alves et al. (2014).

Referente ao número total de estações e a quantas delas monitoravam o MP_{10} e o $MP_{2.5}$, os números apresentados por ambos os estudos são da mesma ordem de grandeza. Relativo ao monitoramento de MP_{10} , apesar de previsto na Resolução CONAMA nº 3/1990, nem todas as estações em cada estado o realizavam (Tabela 1.2). Já o monitoramento do $MP_{2.5}$, só acontece nos estados de São Paulo e do Rio de Janeiro. Espera-se que o número de estados o monitorando aumente, já que a Resolução CONAMA nº 491/2018 passou a prever o seu monitoramento em todo território nacional.

Tabela 1.2: *Número total de estações de monitoramento da qualidade de ar e número de estações que monitoravam o MP_{10} e o $MP_{2.5}$ nas unidades federativas que operavam redes de monitoramento da qualidade do ar no Brasil, de acordo com IEMA (2014) e Alves et al. (2014).*

Estado	OEMA	IEMA (2014)			Alves et al. (2014)		
		Total	MP_{10}	$MP_{2.5}$	Total	MP_{10}	$MP_{2.5}$
DF	IBRAM	5	1	0	4	4	0
ES	IEMA-ES	8	8	0	8	8	0
MG	FEAM	9	9	1	20	7	0
PR	IAP	12	6	0	13	6	0
RJ	INEA	108	67	24	80	54	18
BA	empresas privadas	21	10	0	16	10	0
RS	FEPAM	24	20	0	20	17	0
SP	CETESB	83	53	8	86	78	14
SE	ADEMA	1	0	0	1	1	0
GO	SEMARH				3	0	0
MT	SEMA				1	0	0
		271	174	33	248	185	32

IBRAM: Instituto Brasília Ambiental; IEMA-ES: Instituto Estadual de Meio Ambiente e Recursos Hídricos do Espírito Santo; FEAM: Fundação Estadual do Meio Ambiente; IAP: Instituto Ambiental do Paraná; INEA: Instituto Estadual do Ambiente; FEPAM: Fundação Estadual de Proteção Ambiental; CETESB: Companhia Ambiental do Estado de São Paulo; ADEMA: Administração Estadual do Meio Ambiente de Sergipe; SEMARH: Secretaria de Estado do Meio Ambiente e dos Recursos Hídricos; SEMA: Secretaria de Estado do Meio Ambiente

A Tabela 1.3, reproduzida de Alves et al. (2014), apresenta uma comparação entre o Brasil, o estado de São Paulo, o estado do Rio de Janeiro, os Estados Unidos e a Europa, para o número de estações por 1.000 km^2 e por 100.000 habitantes. Os resultados mostram que o estado de São Paulo e os Estados Unidos apresentam números similares de estações por 1.000 km^2 e o mesmo pôde ser observado para o estado do Rio de Janeiro e a Europa. O número de estações por 1.000 km^2 do Brasil é o que mais se distancia dos demais. No que diz respeito ao número de estações por 100.000 habitantes, Estados Unidos e Europa têm no mínimo 11, 7 e 3 vezes o número de estações do Brasil, São Paulo e Rio de Janeiro, respectivamente, evidenciando o quanto a rede de monitoramento da qualidade do ar no Brasil, de uma forma geral, ainda precisa evoluir.

Tabela 1.3: *Número de estações por 1.000 km² e por 100.000 habitantes*

País/ Estado	Número de estações	Área do território	População total	Estações/1.000 km ²	Estações/100.000 hab.
Brasil	252	8.515.767	190.732.694	0,03	1,3
São Paulo	86	248.222	41.262.199	0,35	2
Rio de Janeiro	80	43.78	15.989.929	1,8	5
EUA	5.000*	9.826.675	318.154.828	0,5	16
Europa	7.500	4.234.000	505.665.739	1,7	14,8

Fonte: *Alves et al. (2014)*; *número de estações ativas

1.6 Monitoramento dos aerossóis via sensoriamento remoto

Além do Brasil, outros países em desenvolvimento carecem de redes de monitoramento da qualidade do ar (Fajersztajn et al., 2014; WHO, 2016). Diante desse cenário, a integração dos dados obtidos do monitoramento via sensoriamento remoto por satélites, com os dados obtidos da rede de monitoramento na superfície, pode contribuir para uma melhor caracterização da variabilidade espaço-temporal dos aerossóis, em uma escala local ou regional (de Almeida Castanho, 2005).

Um exemplo dessa integração de dados é a aplicação IDEA (da sigla em inglês *Infusing satellite data into environmental applications*) desenvolvida por meio de uma parceria entre a Agência Espacial Norte Americana (NASA), a Agência de Proteção Ambiental dos Estados Unidos (USEPA) e a Administração Nacional Oceânica e Atmosférica (NOAA). A ideia é que a USEPA e a NOAA incluam em suas análises os dados de satélite disponibilizados pela NASA, com o objetivo de melhorar a avaliação, o gerenciamento e a previsão da qualidade do ar. Mapas de parâmetros obtidos dos sensores MODIS (*Moderate Resolution Imaging Spectroradiometer*) e GASP (*GOES Aerosol and Smoke Product*) são disponibilizados diariamente em <https://www.star.nesdis.noaa.gov/smcd/spb/aq/>, auxiliando pesquisadores e especialistas que trabalham em agências estaduais ou locais de gerenciamento de qualidade do ar, no sentido de fornecerem informações sobre a distribuição e o transporte dos aerossóis sobre os Estados Unidos (Al-Saadi et al., 2005; Hoff and Christopher, 2009).

A POA (Profundidade Óptica do Aerossol), uma propriedade óptica do aerossol que mede o quanto a radiação foi atenuada ao atravessar a coluna atmosférica, é uma das medidas obtidas do monitoramento da qualidade do ar via sensoriamento remoto por satélite (Figura 1.4). A atenuação ocorre devido à interação (absorção e espalhamento) da radiação com as componentes da atmosfera (gases e partículas) e com a superfície (Wallace and Hobbs, 2006; Brasseur and Prinn, 2003). A POA assume valores positivos e tem sido utilizada como variável explicativa em modelos de predição da concentração de MP_{2,5} na superfície em diferentes regiões do mundo (Chu et al., 2016), e também foi utilizada nesta tese.

Figura 1.4: Ilustração da interação (absorção e espalhamento) da radiação solar com os componentes da atmosfera (gases e partículas) e com a superfície.

Pelo menos dez sensores a bordo de satélites mediam a POA em 2019 (Sogacheva et al., 2019). Os dados de POA do sensor MODIS são os mais utilizados, como demonstrado por Chu et al. (2016), que revisaram os estudos que utilizaram algum modelo para a predição do $MP_{2.5}$ a partir de medidas de POA obtidas de satélite. Cento e dezesseis estudos foram incluídos na revisão realizada por Chu et al. (2016), sendo que 79% deles utilizaram dados de POA do sensor MODIS: 67% utilizaram dados de POA provenientes apenas do sensor MODIS e 12% utilizaram dados de POA tanto do sensor MODIS quanto do sensor MISR (*Multi-angle Imaging SpectroRadiometer*).

Sensores MODIS estão a bordo dos satélites Terra e Aqua, e orbitam a uma altitude de aproximadamente 700 km, varrendo uma faixa de 2330 km (horizontal) por 10 km (vertical). O Terra e o Aqua estão em órbita desde dezembro de 1999 e maio de 2002, e cruzam a linha do Equador aproximadamente às 10:30 e às 13:30 (no horário local), respectivamente.

A POA também é monitorada em estações localizadas na superfície. A AERONET (*Aerosol RObotic NETwork*) é uma rede mundial de radiômetros solares de monitoramento de aerossóis (<https://aeronet.gsfc.nasa.gov>; Holben et al. (1998, 2001)), que em 2019 contava com cerca de 450 estações localizadas em todo o mundo (Sogacheva et al., 2019). Por serem mais acurados, os dados da rede AERONET são utilizados tanto na validação dos dados de satélite quanto na caracterização dos aerossóis em todo o mundo (Levy et al., 2007, 2010). No processo de validação, os dados obtidos do sensor no satélite são pareados, no tempo e no espaço, com os dados da AERONET, de forma que possam ser comparados. Nesse processo, além de se quantificar as incertezas dos dados obtidos via satélite, observam-se oportunidades para melhorias nos algoritmos que são utilizados na obtenção desses dados

(Levy et al., 2007).

A resolução espacial dos dados de POA obtidos do MODIS é um importante fator limitante para os estudos em áreas urbanas (escala local) (de Almeida Castanho, 2005)). No período em que esta tese foi escrita, a versão dos dados de POA do MODIS (versão 6) era distribuída nas resoluções espaciais de 10 km^2 (Levy et al., 2013; Sayer et al., 2013), 3 km^2 (Remer et al., 2013; Munchak et al., 2013; Gupta et al., 2018) e 1 km^2 (Lyapustin et al., 2018).

A resolução espacial de 10 km^2 é considerada satisfatória para o monitoramento dos aerossóis em escala regional e global, enquanto que a resolução de 3 km^2 foi criada para atender a demanda da comunidade científica por um produto com maior resolução espacial, porém ainda é inadequada para estudos de pequena escala, como os realizados em áreas urbanas (Chudnovsky et al., 2014a). Atualmente, os dados de POA obtidos do MODIS com melhor resolução espacial são os fornecidos pelo algoritmo *Multiangle Implementation of Atmospheric Correction* (MAIAC), que distribui esses dados com uma resolução espacial igual a 1 km^2 (Lyapustin et al., 2018), mais adequados para estudos em escala local (Chudnovsky et al., 2014a).

1.7 Modelo Linear Misto (MLM) e a *Metodologia de Harvard* para a predição da concentração de $MP_{2.5}$ na superfície

Pesquisadores da Faculdade de Saúde Pública de Harvard desenvolveram uma metodologia em quatro estágios para estimar a concentração diária de $MP_{2.5}$ e/ou MP_{10} em cada ponto da grade definida sobre a região de *New England*, que inclui os estados de *Maine*, *Vermont*, *New Hampshire*, *Massachusetts*, *Rhode Island* e *Connecticut*, localizados na região Nordeste dos Estados Unidos (Lee et al., 2011; Kloog et al., 2011; Chudnovsky et al., 2014b). Essa metodologia permite realizar a predição das concentrações diárias de $MP_{2.5}$ e/ou MP_{10} em locais sem estação de monitoramento da qualidade do ar, possibilitando o estudo de desfechos associados a períodos de curta e de longa exposição ao $MP_{2.5}$ e/ou MP_{10} . Ela já foi utilizada para realizar a predição das concentrações $MP_{2.5}$ e/ou MP_{10} em outras regiões dos Estados Unidos (Kloog et al., 2012; Lee et al., 2016), na Itália (Stafoggia et al., 2017), na Suíça (de Hoogh et al., 2018), na China (Xiao et al., 2017), na Cidade do México (Just et al., 2015) e em Israel (Kloog et al., 2015), por exemplo. A seguir descrevemos de forma resumida a *Metodologia de Harvard* que também está esquematizada na Figura 1.5.

O primeiro estágio da metodologia é baseado em dados de POA obtidos de satélite e em uma metodologia estatística chamada Modelo Linear Misto (MLM, Laird and Ware (1982)). Os dados de POA têm sido utilizados como preditores das concentrações de $MP_{2.5}$ em diversos estudos em todo o mundo, pois no geral observa-se uma boa correlação entre

os valores de POA e as concentrações de $MP_{2.5}$ medidos na superfície (Chu et al., 2016). Nesse estágio é ajustado um MLM que estabelece uma relação linear entre as concentrações de $MP_{2.5}$ (variável resposta) com os dados de POA e com as demais variáveis explicativas (variáveis meteorológicas e de uso e ocupação do solo) incluídas no modelo.

Na equação (1.1) temos o modelo do primeiro estágio da *Metodologia de Harvard* ajustado aos dados coletados nesta tese. Nesse estágio são utilizadas apenas as células que possuem tanto a informação de $MP_{2.5}$ quanto a informação de POA e das demais variáveis explicativas, e são incluídos no modelo intercepto e inclinação aleatórias para a POA (u_t e v_t , respectivamente), com a finalidade de contemplar a variação diária da relação das concentrações de $MP_{2.5}$ com a POA devido às variáveis meteorológicas.

$$MP_{it} = \alpha + u_t + (\beta_1 + v_t)POA_{it} + \beta_2 Temperatura_{it} + \beta_3 UmidadeRelativa_{it} + \beta_4 ACLP_{it} + \beta_5 Precipitacao_{it} + \beta_6 VelocidadedoVento_{it} + \beta_7 DensidadeViaria_i + \epsilon_{it} \quad (1.1)$$

em que MP_{it} corresponde à média diária da concentração de $MP_{2.5}$ no dia t na célula i ; α e u_t são respectivamente, os interceptos fixo e aleatório; POA_{it} é o valor de POA observado na célula i no dia t ; β_1 e v_t são as inclinações fixa e aleatória, respectivamente; $Temperatura_{it}$, $UmidadeRelativa_{it}$, $ACLP_{it}$, $Precipitacao_{it}$ e $VelocidadedoVento_{it}$ são os valores de temperatura da superfície, umidade relativa, altura da camada limite planetária, precipitação acumulada e velocidade do vento correspondentes à célula i no dia t ; $DensidadeViaria_i$ é o valor da densidade viária correspondentes à célula i ; supõe-se $\epsilon_{it} \sim N(0, \sigma^2)$ e $(u_t, v_t) \sim N[(0, 0), \Sigma]$, mutuamente independentes. Mais detalhes sobre essa etapa da *Metodologia de Harvard* são apresentados na seção 3.3.1.

No estágio 2, utilizando as estimativas dos parâmetros do modelo obtidas no estágio 1, são preditas as concentrações diárias de $MP_{2.5}$ para as células nas quais não existem estações de monitoramento mas existem os dados de POA e das demais variáveis explicativas.

No estágio 3, por meio das estimativas obtidas no estágio 2 e utilizando uma suavização (*spline*) da latitude e da longitude, as concentrações de $MP_{2.5}$ são preditas para as células que não possuem informação de POA, e dessa forma a predição fica completa para toda a grade.

Finalmente, no estágio 4 as concentrações diárias de $MP_{2.5}$ preditas podem ser associadas a uma residência específica, por meio de sua latitude e longitude.

Figura 1.5: Esquematização das etapas 1-3 da Metodologia de Harvard.

A aplicação mais comum do MLM é nas análises de dados de estudos com medidas repetidas, ou seja, quando mais de uma medida/observação é realizada na mesma unidade amostral. Os estudos longitudinais são exemplos de estudos com medidas repetidas. Nesses estudos, como uma mesma unidade amostral é avaliada/medida ao longo do tempo, espera-se que tais medidas sejam correlacionadas. Levar em conta essa correlação na estimação dos parâmetros do modelo melhora a precisão das estimativas obtidas. Mais detalhes sobre o MLM podem ser encontrados em [Laird and Ware \(1982\)](#) e [Singer et al. \(2018\)](#), por exemplo.

No problema que estamos estudando na tese, existem dois tipos de correlação que devem ser consideradas na estimação dos parâmetros do MLM: a espacial e a temporal. A inclusão dos efeitos aleatórios u_t e v_t , como proposto no MLM do primeiro estágio da *Metodologia de Harvard* (equação 1.1), contempla a correlação existente entre as concentrações observadas em um mesmo dia (correlação espacial). No entanto, o MLM proposto no primeiro estágio da *Metodologia de Harvard* ignora a correlação existente entre as concentrações observadas em uma mesma estação de monitoramento ao longo do tempo (correlação temporal). Uma das contribuições da tese, que será detalhada na seção 3.3.3.1, foi a proposta de um MLM no primeiro estágio que considera essas duas fontes de correlação.

Além disso, pensando no contexto da RMSP, onde as emissões veiculares correspondem a 40% da composição do MP_{2.5} e/ou do MP₁₀ ([Martins Pereira et al., 2017](#)), propusemos MLMs que além das variáveis meteorológicas e de uso e ocupação do solo, consideraram variáveis que não costumam ser utilizadas na *Metodologia de Harvard*: as variáveis indicadoras da ocorrência de feriado e dos dias da semana, bem como a interação entre elas; e termos específicos para capturar a sazonalidade das concentrações de MP_{2.5} ao longo do tempo.

Mais detalhes sobre essa estratégia são apresentados na seção 3.3.3.

2. Objetivos

2.1 Objetivo Geral

Estimar as concentrações diárias de $MP_{2.5}$, em uma escala local na RMSP, no período de janeiro 2012 a dezembro de 2017.

2.2 Objetivos Específicos

- calibrar às condições da RMSP, o MLM proposto no primeiro estágio da *Metodologia de Harvard*, para estimar as concentrações de $MP_{2.5}$ no período de janeiro de 2012 e dezembro de 2017;
- estudar a relação entre as concentrações de $MP_{2.5}$ e MP_{10} medidas na superfície com os dados de POA com alta resolução espacial (1 km^2) obtidos do algoritmo *Multiangle Implementation of Atmospheric Correction* (MAIAC) na RMSP;
- propor e avaliar o poder de predição de um MLM que considera as duas fontes de correlação existente nos dados (espacial e a temporal), para estimar as concentrações de $MP_{2.5}$ na RMSP;
- propor e avaliar o poder de predição de um MLM que, além das variáveis meteorológicas e de uso e ocupação do solo, também considera as variáveis feriado e dia da semana, bem como o efeito de interação entre elas, para estimar as concentrações de $MP_{2.5}$ na RMSP;
- propor e avaliar o poder de predição de um MLM que, além das variáveis meteorológicas, das variáveis de uso e ocupação do solo, das variáveis feriado e dia da semana, bem como o efeito de interação entre elas, também considera termos específicos com o intuito de capturar a sazonalidade das concentrações de $MP_{2.5}$ ao longo do tempo, para estimar as concentrações de $MP_{2.5}$ na RMSP;
- além das estimativas pontuais das concentrações de $MP_{2.5}$, apresentar uma medida de incerteza associada às estimativas obtidas nos MLMs citados anteriormente.

3. Métodos

3.1 Período e região do estudo

Como os óbitos incluídos no Projeto MODAU ocorreram majoritariamente nos anos de 2017 e 2018, definiu-se o período do estudo o compreendido entre janeiro de 2012 e dezembro de 2017. Um período de pelo menos cinco anos anteriores ao desfecho costuma ser suficiente para avaliar os efeitos da exposição de curto e de longo prazo ao $MP_{2.5}$ (Brunekreef and Holgate, 2002; Pope et al., 1995; Spix et al., 1998; Andersen et al., 2012; Hrubá et al., 2001; Oudin et al., 2011).

A região do estudo foi a RMSP, que compreende a Cidade de São Paulo e outras 38 cidades (Figura 3.1). É uma região urbana que abrange uma área de 7946 km^2 . Com mais de 19 milhões de habitantes, sendo que 12 milhões vivem na Cidade de São Paulo, a RMSP é a região mais populosa da América Latina (<https://emplasa.sp.gov.br/RMSP>) sendo as emissões veiculares sua principal fonte de poluição do ar (CETESB, 2019a).

Em relação ao clima, existem duas estações principais na RMSP: a chuvosa e a seca. A primeira dura entre outubro e março e a última entre abril e setembro (Andrade et al., 2012a). Devido às condições meteorológicas, o período compreendido entre maio e setembro costuma ser o mais desfavorável para a dispersão de poluentes no estado de São Paulo. Nesse período são frequentes as ultrapassagens do padrão diário de $60 \mu\text{g}/\text{m}^3$ para o $MP_{2.5}$ estipulado pelo Decreto Estadual nº 59.113/2013 (CETESB, 2019a).

Figura 3.1: Localização das estações da rede automática da CETESB na RMSP que monitoravam o $MP_{2.5}$ e/ou o MP_{10} em 2017, da estação AERONET-SP e do Aeroporto Campo de Marte

3.2 Dados primários e seu processamento

Como o objetivo da tese é a construção de um modelo espaço-temporal, construímos um banco de dados visando esse objetivo. Para tal, o banco de dados possui dois identificadores: um espacial e um temporal. O identificador temporal é a data. Já o identificador espacial foi construído da seguinte forma: sobre a RMSP foi definida uma grade composta por células de 1 km^2 (resolução do dado de satélite), como esquematizado na Figura 3.2; a primeira célula do canto superior esquerdo recebeu o identificador “1”, a célula à sua direita recebeu o identificador “2” e assim por diante, até a última célula no canto inferior direito, que recebeu o identificador “15092”. Assim, cada observação no banco de dados fica identificada de forma única, no tempo e no espaço.

Figura 3.2: Grade com células de 1 km^2 sobre a RMSP

Como os conjuntos de dados utilizados no estudo vieram de várias fontes e em diferentes formatos, para a confecção do banco de dados utilizado nas análises (denominado banco de dados integrado), todos os conjuntos de dados primários passaram por algum tipo de processamento, detalhado nas seções que seguem.

Para cada tipo de dado (concentração de poluentes da CETESB, meteorológicos, dados de satélite, mapa de uso e ocupação do solo) o resultado do processamento é um banco de dados em formato padrão, como ilustrado na Tabela 3.1, que vamos chamar de banco de dados padronizado.

Tabela 3.1: *Esquematização de um banco de dados padronizado.*

Data	número da célula	V_1	V_2	...	V_N
01/01/2012	1				
01/01/2012	2				
...	...				
01/01/2012	15092				
...	...				
31/12/2012	1				
31/12/2012	2				
...	...				
31/12/2012	15092				
...	...				
01/01/2017	1				
01/01/2017	2				
...	...				
01/01/2017	15092				
...	...				
31/12/2017	1				
31/12/2017	2				
...	...				
31/12/2017	15092				

V_k : k -ésima variável do banco de dados

A integração dos bancos de dados padronizados tem como resultado o banco de dados integrado, que foi utilizado nas análises. No total, foram construídos quatro bancos de dados integrados, sendo dois referentes aos dados do momento da passagem dos satélites (banco integrado horário Aqua, banco integrado horário Terra) e os demais referentes aos dados diurnos e diários (banco integrado diurno e banco integrado diário).

A seguir descrevemos o processamento realizado nos dados primários que deram origem aos dados que foram utilizados no ajuste dos modelos da tese. Como exemplificado na equação (1.1), a concentração de $MP_{2.5}$ obtida das estações de monitoramento da qualidade do ar foi a variável resposta, enquanto os dados de POA, os dados meteorológicos (preditores espaço-temporais) e as variáveis de uso e ocupação do solo (preditores espaciais) foram as variáveis explicativas.

Os bancos de dados integrados diurno e diário serão disponibilizados em um repositório de acesso livre (<https://repositorio.uspdigital.usp.br/>) assim que os resultados referentes aos modelos desenvolvidos na tese forem publicados.

3.2.1 Concentrações de material particulado

No estado de São Paulo, o órgão ambiental responsável pelo monitoramento da qualidade do ar é a CETESB (Companhia Ambiental do Estado de São Paulo), que opera redes de

monitoramento automática e manual, compostas por estações fixas e móveis. Em 2017, na RMSP, a rede automática tinha 29 estações fixas e uma estação móvel (CETESB, 2018). São monitoradas as concentrações de $MP_{2.5}$, MP_{10} , CO, NO, NO_2 , NO_X , SO_2 e O_3 , além de algumas variáveis meteorológicas (Temperatura, Umidade Relativa, Direção e Velocidade do Vento, por exemplo), não necessariamente em todas as estações. As concentrações dos poluentes bem como os dados meteorológicos podem ser obtidos no sistema QUALAR (<http://ar.cetesb.sp.gov.br/qualar/>). A resolução temporal dos dados disponibilizados pela CETESB é a horária.

A Figura 3.3 apresenta todo o processamento realizado nos dados obtidos da CETESB. O primeiro processamento foi a integração dos conjuntos de dados primários, que foram disponibilizados em arquivos separados por variável e ano. Esse processamento deu origem ao banco de dados com as concentrações dos poluentes e ao banco de dados com as variáveis meteorológicas. Note que esses dois bancos de dados têm resolução temporal horária.

Neste estudo foram consideradas as concentrações de $MP_{2.5}$ e MP_{10} ($\mu g/m^3$) obtidas da rede automática. Foram calculadas as médias diurnas (média das concentrações observadas entre 6h-18h, para um determinado dia) e diárias (média das concentrações observadas nas 24h do dia), que por sua vez foram consideradas válidas quando calculadas com pelo menos 2/3 de observações horárias válidas, definidas pela CETESB (CETESB, 2019a).

Também foram consideradas as concentrações de $MP_{2.5}$ e MP_{10} observadas na hora da passagem dos satélites Aqua e Terra sobre a RMSP. Assim, ao final desse processamento temos cinco bancos de dados: um banco com as concentrações observadas no momento da passagem do satélite Terra, um banco com as concentrações observadas no momento da passagem do satélite Aqua, um banco com as concentrações diurnas, um banco com as concentrações diárias e o banco horário com as variáveis meteorológicas. Com exceção do último, os demais deram origem aos bancos de dados padronizados utilizados na construção do banco de dados integrado final.

Figura 3.3: Esquematização do processamento dos dados obtidos da CETESB.

A Tabela 3.2 apresenta a evolução do número de estações com dados válidos, definidos pela CETESB, que monitoraram o $MP_{2.5}$ na RMSP no período do estudo. Observa-se que o

número de estações com dados válidos de $MP_{2.5}$ aumenta no decorrer do período do estudo sendo que apenas 5 e 14 estações apresentaram dados válidos de $MP_{2.5}$ na RMSP em 2012 e 2017, respectivamente.

Tabela 3.2: *Estações com dados válidos de $MP_{2.5}$ na RMSP no período do estudo*

Estação	2012	2013	2014	2015	2016	2017
Congonhas	X	X	X	X	X	X
USP-IPEN	X	X	X	X	X	X
Marginal Tietê-Ponte dos Remedios	X	X	X	X	X	X
Parelheiros	X	X	X	X	X	X
Pinheiros	X	X	X	X		X
Ibirapuera			X	X	X	X
São Bernardo-Centro			X	X	X	X
Guarulhos-Pimentas				X	X	X
Itaim Paulista				X	X	X
Pq. D. Pedro II					X	X
Pico do Jaraguá					X	X
Guarulhos-Paco Municipal						X
Santana						X
Osasco						X
Total	5	5	7	9	10	14

Em seus relatórios anuais de qualidade do ar, a CETESB informa que na RMSP a concentração de $MP_{2.5}$ corresponde a cerca de 60% da concentração do MP_{10} , ou seja, a razão $MP_{2.5}/MP_{10}$ é aproximadamente igual a 0,6. Esse resultado é baseado em dados de um estudo que analisou as primeiras medições de concentrações de $MP_{2.5}$ obtidos na RMSP. A cada seis dias, por períodos contínuos de 24 horas, foram coletadas amostras das frações <2.5 e 2.5-10 do MP (método dicotômico) nas estações de monitoramento manual Cerqueira César, Pinheiros, Ibirapuera e São Caetano do Sul no período de 2000-2006. Mais detalhes sobre esse estudo podem ser encontradas em [CETESB \(2008\)](#).

Como a razão $MP_{2.5}/MP_{10}$ na RMSP divulgada pela CETESB foi estimada com base em dados do período 2000-2006, decidimos analisar os dados das estações de monitoramento manual referentes ao período 2000-2017. No entanto, os dados referentes ao $MP_{2.5}$ e ao MP_{10} amostrados com o método dicotômico, só estavam disponíveis no sistema QUALAR até o ano de 2014 (último acesso em setembro de 2020). Os resultados dessa análise, apresentados no Apêndice A, mostram uma tendência de queda nos valores da razão $MP_{2.5}/MP_{10}$ na RMSP, sendo que no período 2012-2014, em média, a razão $MP_{2.5}/MP_{10}$ está mais próxima de 0,5 do que de 0,6.

Com base nesses resultados, consideramos que a concentração de $MP_{2.5}$ na RMSP corresponde a cerca de 50% da concentração do MP_{10} . Deste modo, utilizamos as concentrações de $MP_{2.5}$ nas estações em que estavam disponíveis, e nas estações que mediam o MP_{10} consideramos que $MP_{2.5} \sim 0,5 \times MP_{10}$. Como o número de estações medindo o MP_{10} na RMSP

é maior do que o número de estações medindo o $MP_{2.5}$, ao utilizar essa aproximação, aumentamos a cobertura espacial do $MP_{2.5}$ na RMSP. Esse tipo de aproximação costuma ser utilizada, por exemplo, em modelos de predição de concentrações de poluentes desenvolvidos pela OMS (WHO, 2016).

3.2.2 Preditores espaço-temporais do $MP_{2.5}$

3.2.2.1 Dados de sensoriamento remoto

Neste estudo, foram utilizados dados de POA obtidos de sensores em satélites e em superfície.

Referente aos dados de satélite, foram utilizados dados de POA com resolução espacial de 1 km^2 fornecidos pelo algoritmo MAIAC, no comprimento de onda de $0,55\text{ }\mu\text{m}$, extraídos da versão 6 dos dados do sensor MODIS que está a bordo dos satélites Aqua e Terra.

A validação dos dados do MAIAC, aos quais vamos nos referir como MODIS-MAIAC, mostrou que a precisão dos dados de POA obtidos sobre superfícies com vegetação escura é comparável a do produto com resolução de 10 km^2 , além de apresentar maior precisão sobre superfícies brilhantes (por exemplo, desertos ou áreas urbanas) (Lyapustin et al., 2011a).

Os dados de POA do MODIS-MAIAC já foram validados para a América do Sul (Martins et al., 2017), e nessa região, já foram utilizados em estudos na região da floresta Amazônica (Alden et al., 2016; Lopes et al., 2016; Saleska et al., 2016; Maeda et al., 2017; Wagner et al., 2017). Até onde é de nosso conhecimento, esta tese foi a primeira a utilizar esses dados na RMSP.

No período do estudo, a passagem do satélite Terra sobre a RMSP ocorreu entre 9h e 11h30, enquanto que a passagem do satélite Aqua ocorreu entre 12h45 e 15h05 (horário local). Os dados do MAIAC estão disponíveis em <https://earthdata.nasa.gov/> e são distribuídos no formato HDF (*Hierarchical Data Format*), sendo um arquivo para cada dia do período de estudo. Como nos arquivos HDF são disponibilizadas outras informações além da POA, eles precisam passar por um processamento com o intuito de extrairmos a informação de nosso interesse, no caso a POA. Além disso, a POA é fornecida na projeção sinusoidal. A Figura 3.4 resume o processamento realizado nos dados do MODIS-MAIAC.

Figura 3.4: Esquematização do processamento realizado nos dados do MODIS-MAIAC.

Após a solicitação e o *download* dos dados, a ferramenta *HDF-EOS To GeoTIFF Conversion* (HEG, disponível em <https://newsroom.gsfc.nasa.gov/sdptoolkit/HEG/>

HEGHome.html) foi utilizada para realizar a reprojeção do sistema de coordenadas (da projeção sinusoidal para a projeção geográfica: *datum* WGS84 e sistema de coordenadas UTM) e conversão de formatos (de HDF para *geoTIFF*) dos arquivos HDF. Após essa etapa, os dados referentes a cada satélite foram organizados nos respectivos bancos de dados padronizados, utilizados na integração que deu origem ao banco de dados final.

A Figura 3.5 apresenta o resultado do processamento dos dados de POA do MODIS-MAIAC referentes a uma passagem do satélite Aqua no dia 29/05/2017. A POA é um número adimensional e está diretamente relacionado com a quantidade de aerossol na coluna atmosférica: um valor de 0,01 indicaria uma atmosfera extremamente limpa enquanto um valor de 0,4 indicaria uma atmosfera poluída (NOAA-ESRL, 2020).

Figura 3.5: Mapa da POA MAIAC obtido a partir dos dados do satélite Aqua no dia 29/05/2015 sobre a RMSP.

Apesar de a resolução temporal dos dados ser diária, não necessariamente os dados estarão disponíveis todos os dias e em toda a região de estudo. A presença de nuvens no momento de passagem dos satélites é um fator limitante para a aquisição dos dados de POA, para os instrumentos em satélites ou na superfície. A presença de nuvens “contamina os dados”, no sentido de aumentar os valores da POA. Por isso, uma das etapas dos algoritmos de aquisição dos dados de POA é a remoção dos *pixels* nos quais foram detectados nuvens, e essa remoção gera dados faltantes. A aquisição de dados de POA via satélite também fica prejudicada sobre superfícies brilhantes/refletoras, como os desertos, as superfícies cobertas por neve e regiões com pouca ou nenhuma cobertura vegetal (por exemplo, áreas urbanas) (de Almeida Castanho, 2005).

Também utilizamos dados da rede AERONET das estações São Paulo e GSFC (NASA Goddard Space Flight Center), para o período compreendido entre 2000 e 2011 (dados *Level 2*). Esses dados foram utilizados na comparação de algumas propriedades dos aerossóis dessas estações, com o intuito de verificar se o modelo de aerossol utilizado pelo algoritmo MAIAC

na aquisição dos dados de POA sobre a RMSP é adequado. Mais detalhes são apresentados na seção 3.3.2.3.

3.2.2.2 Dados meteorológicos

Os dados meteorológicos utilizados no estudo foram fornecidos pelas seguintes fontes: Centro de Gerenciamento de Emergências da Cidade de São Paulo (CGE), CETESB, Estação Meteorológica do IAG-USP e Departamento de Águas e Energia Elétrica (DAEE). Foram disponibilizados dados referentes às variáveis temperatura da superfície ($^{\circ}\text{C}$), umidade relativa do ar (%), precipitação (mm), direção do vento (graus) e velocidade do vento (m/s). Com exceção das estações da CETESB, todas as estações medem precipitação. As estações do DAEE medem exclusivamente dados pluviométricos/fluviométricos. Para ilustração, as Figuras 3.6 e 3.7 apresentam a localização das estações meteorológicas e pluviométricas utilizadas no estudo para o ano de 2015.

Figura 3.6: Localização das estações da CETESB, do CGE e IAG-USP que mediram as variáveis meteorológicas consideradas no estudo em 2015.

Figura 3.7: Localização das estações do DAEE que mediram precipitação em 2015.

A resolução temporal dos dados disponibilizados pelo IAG-USP, CGE, DAEE e CETESB é a horária, 10 minutos, diária e horária, respectivamente. A Figura 3.8 apresenta todo o processamento realizado nos dados obtidos do CGE, IAG-USP e DAEE.

Figura 3.8: Esquematização do processamento dos dados meteorológicos.

De forma similar ao processamento realizado nos dados da CETESB, os conjuntos de dados primários de cada fonte foram integrados em um arquivo único: os dados do CGE e do DAEE foram disponibilizados em arquivos separados por estação enquanto os dados do IAG foram disponibilizados em arquivos separados por variável. Após esse procedimento, a integração dos bancos do CGE, CETESB e IAG-USP deu origem ao banco de dados unificado com as variáveis meteorológicas, com resolução temporal horária. A partir desse banco, similarmente ao procedimento realizado com os dados da CETESB, foram construídos os bancos com os dados meteorológicos diurnos e diários, bem como os bancos de dados horários com os dados meteorológicos nos momentos de passagem dos satélites. Os dados do DAEE foram incorporados apenas ao banco de dados diário. Assim, ao final desse processamento temos quatro banco de dados: um banco com as variáveis meteorológicas observadas no momento da passagem do satélite Terra, um banco com as variáveis meteorológicas observadas no momento da passagem do satélite Aqua, um banco com as variáveis meteorológicas diurnas e um banco com as variáveis meteorológicas diárias.

Para as variáveis temperatura e umidade relativa foram calculadas as médias diurnas e diárias. Para a precipitação foram considerados os valores acumulados diários, inclusive para os bancos de dados horários e diurno. Para a velocidade do vento, que é uma grandeza vetorial (definida por um valor numérico, direção e sentido), as médias diurnas e diárias foram calculadas com base nas médias das componentes u e v da decomposição dos valores horários.

Após a construção dos bancos de dados com as variáveis meteorológicas (banco horário Terra, banco horário Aqua, banco diurno e banco diário), para cada dia e cada variável, o método do vizinho mais próximo (Hartmann et al., 2018) foi utilizado para a interpolação dos dados meteorológicos na grade esquematizada na Figura 3.2. Esse método atribui às células em que não existem estações meteorológicas os valores das variáveis meteorológicas

da célula mais próxima e cujos valores estavam disponíveis (Figura 3.9).

Ao final desse processo, os dados foram armazenados nos respectivos bancos de dados padronizados utilizados na construção do banco de dados integrado final. Utilizamos o *software* R (www.r-project.org) versão 3.2.1 no processamento dos dados meteorológicos.

Figura 3.9: Resultado do método do vizinho mais próximo para a interpolação dos dados de temperatura na grade esquematizada na Figura 3.2 no dia 27/02/2017.

3.2.2.2.1 Altura da Camada Limite Planetária

A camada limite planetária (CLP), também conhecida como baixa troposfera, se estende até cerca de 1-2 km acima da superfície terrestre, podendo sua altura variar ao longo do dia e de acordo com o tipo de superfície e as condições meteorológicas. A CLP é extremamente afetada pelas dinâmicas que ocorrem na superfície terrestre, pois é onde ocorre a maioria das emissões de gases e vapores, seja por atividades antropogênicas ou biogênicas (Brasseur et al., 1999; Seinfeld and Pandis, 2006). Consequentemente, a altura da CLP (ACLP), além das condições meteorológicas, pode favorecer ou não a concentração/dispersão dos poluentes atmosféricos.

A ACLP foi estimada a partir de dados de sondagens atmosférica que são realizadas diariamente às 00 e 12 UTC (do inglês *Coordinated Universal Time*) em estações localizadas em todo o planeta. Utilizamos os dados das sondagens realizadas pela Força Aérea Brasileira (FAB) no Aeroporto Campo de Marte, na Cidade de São Paulo (Figura 3.1). Esses dados estão disponíveis em <http://weather.uwyo.edu/upperair/sounding.html>.

Sánchez et al. (2020), utilizando diferentes métodos para a estimativa da ACLP, estudaram as propriedades dinâmicas e termodinâmicas da CLP na RMSP. Os autores mostraram que os métodos da Parcela e do Gradiente da Temperatura do Ar apresentaram melhor desempenho para estimar a ACLP nos períodos da manhã e da tarde, respectivamente. A Figura 3.10 apresenta todo o processamento realizado nos dados de sondagem atmosférica.

Figura 3.10: Esquemática do processamento dos dados de sondagem atmosférica.

As estimativas da ACLP no período da manhã foram obtidas por meio dos perfis de temperatura potencial virtual obtidos das sondagens realizadas às 12 UTC (9h no horário local) utilizando o método da Parcela. Nesse método, a ACLP é definida como a altura em que a temperatura potencial virtual é a mesma que a temperatura potencial virtual de superfície. A ACLP no período da tarde foi estimada pela Altura da Camada de Mistura Residual (ACMR), que é uma boa aproximação da ACLP máxima diurna, que costuma ocorrer no período da tarde (Sánchez et al., 2020). A ACMR foi estimada por meio dos perfis de temperatura do ar obtidos das sondagens realizadas às 00 UTC (21h no horário local) utilizando o método do Gradiente da Temperatura do Ar. Nesse método, a ACMR é definida como a altura em que o gradiente (derivada) de temperatura do ar é maior ou igual a zero.

A Figura 3.11 apresenta exemplos dos perfis de temperatura obtidos a partir dos dados de sondagem atmosférica, nos quais foram aplicados os métodos de estimação da ACLP. Mais detalhes sobre esses métodos podem ser encontrados em Stull (2012) e Seidel et al. (2010), por exemplo.

Figura 3.11: Exemplo de perfis de temperatura obtidos a partir dos dados de sondagem atmosférica do dia 08/01/2012. No eixo y temos a altura na qual a medida foi realizada e no eixo x temos a temperatura potencial virtual (painel esquerdo) e a temperatura do ar (painel direito). As linhas horizontais vermelhas indicam as estimativas da ACLP (painel esquerdo) e da ACMR (painel direito) obtidas pelos métodos da Parcela e da temperatura do ar, respectivamente.

Vale salientar que nem sempre é possível obter a ACLP a partir de dados de sondagens atmosférica. Os dados das sondagens podem estar ausentes devido a problemas técnicos durante o lançamento do balão, a problemas durante a aquisição dos dados ou quando o método de estimativa da ACLP não consegue detectar inversões nos perfis de temperatura. As reduções no tamanho da amostra associadas a não obtenção das estimativas da ACLP foram de 53% e 38%, em média, para os períodos da tarde e da manhã, respectivamente. A frequência de dados faltantes foi maior no período da tarde, porque nesse período a ACLP foi estimada indiretamente pela ACMR, que não é observada todos os dias.

Como estamos utilizando dados de apenas um ponto de lançamento de sondagem atmosférica na RMSP, uma limitação dessa metodologia é a que estamos assumindo que a ACLP é homogênea em toda a RMSP, o que pode não ser verdade. Para contornar essa limitação, tentamos utilizar dados do *Climate Forecast System Reanalysis version 2* (CFSRv2) desenvolvido pelo *National Centers for Environmental Prediction* (NCEP). Trata-se de um modelo climático para previsão sazonal, que leva em conta a interação entre as componentes do sistema oceano-atmosfera-superfície (Saha et al., 2014). Os produtos analisados estão disponíveis nas resoluções horizontais de 0,2, 0,5, 1,0 e 2,5 graus (<https://rda.ucar.edu/>) e nós utilizamos o produto *Planetary Boundary Layer Height* na resolução espacial de 0,20 graus.

Os dados foram disponibilizados no formato NetCDF (*Network Common Data Form*), sendo um arquivo para cada dia do período do estudo. A Figura 3.12 apresenta todo o processamento realizado nos dados obtidos do NCEP. Cada arquivo armazena grades (uma para cada hora do dia) com células de tamanhos irregulares (0,2045 x 0,2044 graus). As grades foram então reprojetadas para grades cujas células possuem 1 km² (dimensão igual às das células da grade dos dados de satélite) e armazenadas no banco de dados padronizado.

As ACLP obtidas por meio de dados de sondagens atmosféricas costumam ser o padrão-ouro em comparações com outras metodologias de estimação de ACLP. Ao compararmos a média das ACLP obtidas do modelo do NCEP no entorno do ponto de lançamento da sonda atmosférica com as ACLP obtidas por meio de dados de sondagens atmosféricas, obtivemos uma concordância muito baixa ($\sim 30\%$), e por esse motivo optamos por utilizar apenas as ACLP obtidas por meio de dados de sondagens atmosféricas.

Figura 3.12: Esquematização do processamento dos dados obtidos do NCEP.

3.2.3 Preditores espaciais do $MP_{2.5}$

Nas aplicações da *Metodologia de Harvard* em outras regiões, variáveis como a altitude, a distância até fontes de emissão de poluentes, informações referentes aos inventários de emissões de poluentes, mapa do uso e ocupação do solo, entre outras, costumam ser utilizadas com preditoras espaciais das concentrações do $MP_{2.5}$. A inclusão desses preditores, bem como a inclusão das demais variáveis, vai depender da disponibilidade dessas informações em cada região de estudo. Para a RMSP, tentamos incluir nos modelos ajustados, a altitude, o mapa de uso e ocupação solo e a densidade viária. Devido às limitações que serão detalhadas a seguir, optamos por incluir apenas a densidade viária.

3.2.3.1 Altitude

Uma variável que costuma ser considerada na *Metodologia de Harvard*, mas que não foi utilizada nos modelos desenvolvidos na tese, é a elevação em relação ao nível do mar (altitude). A inclusão dessa variável nos modelos seria uma tentativa de representar a topografia da região de estudo, fator importante para a dispersão ou concentração de poluentes. No entanto, em médias e pequenas escalas, fatores adicionais relacionados ao terreno são mais relevantes. Por exemplo, a rugosidade inserida no terreno devido à presença de áreas urbanas, o que altera a circulação dos ventos, e consequentemente a dispersão dos poluentes (Vardoulakis et al., 2003). Ou seja, a inclusão da variável altitude nos modelos não representa de forma adequada a complexidade da área urbana que existe na região do estudo.

No caso específico da RMSP encontramos uma limitação que pode ser válida para outras áreas urbanas. Como as vias com grande circulação de veículos estão localizadas às margens de rios/córregos ou sobre regiões onde existiam rios/córregos, que são áreas com menores altitudes, encontramos uma correlação negativa e moderada entre as variáveis densidade viária e altitude: quanto menor a altitude maior a densidade viária. Isso indica que essas duas variáveis trazem a mesma informação para o modelo (são colineares), o que afeta a inferência nos modelos de regressão (Kutner et al., 2005). Por esse motivo optamos por não incluir a altitude nos modelos ajustados na tese.

Figura 3.13: Dispersão entre a densidade viária e a altitude das estações de monitoramento.

3.2.3.2 Uso e ocupação do solo

O mapa de uso e ocupação do solo que seria utilizado neste estudo foi obtido de [Macedo et al. \(2016\)](#) (Figura 3.14). Esses autores construíram um mapa de uso e ocupação do solo para a RMSP utilizando imagens do dia 23/09/2015 do satélite LANDSAT 8 e o método de classificação supervisionada de Bhattacharya ([Moreira, 2005](#)), que identificou seis tipos de solo urbano, quatro tipos de cobertura vegetal, regiões com corpos d'água e solo exposto, totalizando doze tipos de uso e ocupação do solo na RMSP.

O solo urbano 1 compreende regiões que possuem uma mescla entre área construída e área vegetada, e os prédios em grande maioria são de porte baixo e/ou médio. O solo urbano 2 é caracterizado pela alta densidade de construções com estruturas irregulares, com pouca ou nenhuma fração vegetada. A região do solo urbano 3 localiza-se em uma área central da RMSP, apresenta um grande número de prédios com grande extensão horizontal e porte baixo. O tipo urbano 4 apresenta características residenciais, com construções baixas e ruas estreitas. O solo urbano 5 engloba a região central da RMSP, apresenta uma grande densidade de prédios e verticalidade pronunciada, avenidas largas e grande tráfego veicular. Construções esparsas de pequeno porte e grande cobertura de áreas cultivadas compõem o solo urbano 6, que apresenta características de solo suburbano.

Figura 3.14: Mapa de uso e ocupação do solo na RMSP construído por Macedo e Freitas, 2016.

Dentre as 12 classes definidas com base na análise das imagens de satélite, a classe que tem uma maior área é a de vegetação. Entretanto, se considerarmos somente a mancha urbana, o tipo de solo urbano 4 é o predominante (Tabela 3.3).

Tabela 3.3: Uso e ocupação do solo na RMSP - Proporção da área ocupada

Classe	Área Ocupada (%)
Urbano 1	6,31
Urbano 2	0,43
Urbano 3	2,98
Urbano 4	20,04
Urbano 5	1,31
Urbano 6	6,09
Vegetação 1	38,51
Vegetação 2	8,95
Vegetação 3	8,98
Vegetação 4	3,63
Solo Exposto	0,07
Corpos d'água	2,70
Área total	100

Na RMSP foi observado que a maioria das estações de monitoramento está localizada

em solo do tipo urbano (2 no tipo urbano 2, 6 no tipo urbano 4, 17 no tipo urbano 5, 1 no tipo urbano 6 e 2 no tipo vegetação), o que não traz informação para o modelo. Por isso, esse preditor de $MP_{2.5}$ não foi considerado nos modelos ajustados na tese.

3.2.3.3 Emissões veiculares e densidade viária

No Brasil e no mundo, as emissões veiculares são as principais fontes de poluição atmosférica em áreas urbanas (Gupta et al., 2006; WMO, 2012; Andrade et al., 2015). Em apenas 3,2% do território do estado de São Paulo, a RMSP concentra 48% de toda a frota circulante (7.284.990 de veículos) e 47% da população (21,6 milhões de habitantes) do estado. Já a Cidade de São Paulo, com 4.573.532 de veículos, concentra respectivamente 63% e 30% da frota circulante da RMSP e do estado (CETESB, 2019b).

Uma revisão dos trabalhos que estudaram a composição do $MP_{2.5}$ e/ou MP_{10} em São Paulo, realizada por Martins Pereira et al. (2017), mostra que as emissões veiculares, a ressuspensão do solo e as emissões industriais constituem respectivamente cerca de 40%, 30% e 10% desses poluentes. Por isso, variáveis relacionadas às emissões veiculares costumam ser boas preditoras da concentração de $MP_{2.5}$ e/ou MP_{10} na Cidade de São Paulo, como demonstrado no estudo de Habermann and Gouveia (2012).

Utilizando dados do “Relatório Emissões Veiculares no Estado de São Paulo” (CETESB, 2019b), construímos a Figura 3.15 que apresenta a contribuição relativa de cada categoria de veículos na emissão de poluentes na RMSP em 2018. No que se refere ao MP, os caminhões e os ônibus urbanos foram as categorias de veículos que apresentaram as maiores contribuições nas emissões de MP na RMSP, com 51% e 21%, respectivamente.

Figura 3.15: Contribuição relativa de cada categoria de veículos na emissão de poluentes na RMSP em 2018. Fonte: CETESB (2019b)

Apesar de os caminhões pesados e semi pesados geralmente circularem nas rodovias e

horária e resolução espacial igual a 1 km^2 e estão disponíveis em <http://emissoes.energiaeambiente.org.br/>.

Como estamos trabalhando com os dados das estações de monitoramento da RMSP, ao utilizar as emissões fornecidas pelo IEMA, que estão disponíveis apenas para a Cidade de São Paulo, acabaríamos perdendo dados de 12 estações de monitoramento. Com o intuito de utilizar toda a informação disponível referente aos dados de monitoramento, optamos por não utilizar as emissões fornecidas pelo IEMA nos modelos ajustados na tese.

Como Just et al. (2015) e de Hoogh et al. (2018) em estudos realizados na Cidade do México e na Suíça, respectivamente, no presente estudo utilizamos a variável densidade viária como uma aproximação para as emissões veiculares na RMSP. Para cada célula da grade esquematizada na Figura 3.2 calculamos a densidade viária (m/m^2 , metro de via por metro quadrado de área) de vias com alta circulação de veículos utilizando a ferramenta *line density tool* do software *ArcGis* versão 10.3 (Figura 3.18).

Figura 3.18: Densidade viária de vias de alta circulação na RMSP.

O mapa de vias foi obtido do *OpenStreetMap* (OSM, *download* realizado em outubro de 2017; openstreetmap.org). As vias consideradas com alta circulação de veículos foram as vias classificadas como *trunk*, *motorway*, *primary* and *secondary roads* pelo OSM, pois assumimos que essas são as vias equivalentes às classificadas como vias expressa, arterial e coletora pela Companhia de Engenharia de Tráfego (CET, (<http://www.cetesp.com.br/media/45667/portaria2102.pdf>)). A Figura 3.19 apresenta o processamento realizado no mapa de vias obtido do OSM.

Figura 3.19: Esquematização do processamento realizado no mapa de vias obtido do OSM.

Como ilustrado na Figura 3.20 verificamos que a correlação entre a densidade viária e as estimativas de emissões de poluentes fornecidas pelo IEMA é alta/moderada, indicando que é razoável utilizar a densidade viária como uma aproximação das emissões veiculares.

Figura 3.20: Dispersão entre a densidade viária e as estimativas de emissões dos poluentes CO, NOx e MP total fornecidas pelo IEMA, na cidade de São Paulo. Os painéis superiores referem-se às emissões diurnas e os inferiores às emissões diárias. R: coeficiente da correlação de Pearson.

3.3 Análise dos dados

3.3.1 Primeiro estágio da *Metodologia de Harvard* para a predição da concentração de $MP_{2.5}$

Como introduzido na seção 1.7, os modelos ajustados na tese foram baseados em uma metodologia proposta por pesquisadores da Faculdade de Saúde Pública de Harvard.

O primeiro estágio dessa metodologia é baseado em dados de POA obtidos de satélite e em um método estatístico chamado Modelo Linear Misto (MLM, [Laird and Ware \(1982\)](#)). Os dados de POA têm sido utilizados como preditores das concentrações de $MP_{2.5}$ em diversos estudos em todo o mundo, pois no geral observa-se uma boa correlação entre os valores de POA e as concentrações de $MP_{2.5}$ medidas na superfície. Como ambos podem variar devido principalmente às condições meteorológicas, é razoável esperar que a relação $MP_{2.5}$ -POA também varie, e é aqui que entra o MLM.

Considere a situação hipotética em que temos as concentrações de $MP_{2.5}$ de 20 estações de monitoramento ao longo de 4 dias. Suponha que y_{ij} e x_{ij} sejam as concentrações de $MP_{2.5}$ e os valores de POA observados na estação de monitoramento i no dia j , respectivamente, em que $i = 1, 2, \dots, 20$ e $j = 1, 2, 3, 4$. A Figura 3.21 ilustra o ajuste de um MLM, em que obtemos uma relação linear para cada dia, atendendo à suposição de que a relação $MP_{2.5}$ -POA varia diariamente. Mas ao contrário do que possa parecer, não estamos ajustando uma reta para cada dia via regressão linear. A variação das relações ao longo dos dias se dá devido à inclusão dos efeitos aleatórios a_j e b_j que possuem uma distribuição normal bivariada, no caso $(a_j, b_j) \sim N((0, 0)^\top, \Sigma)$. Então no fundo, além dos parâmetros da parte fixa do modelo (α e β) e da variância do erro aleatório ϵ_{ij} (σ^2), o que está sendo estimado são as componentes da matriz de variâncias e covariâncias $\Sigma = \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix}$, em que σ_a^2 e σ_b^2 são as variâncias dos efeitos aleatórios a_j e b_j , respectivamente e σ_{ab} é a covariância entre esses dois efeitos. No contexto do MLM, a_j e b_j representam a diferença entre o intercepto e a inclinação da relação ($MP_{2.5}$ -POA) obtida no dia j e o intercepto e a inclinação da relação ($MP_{2.5}$ -POA) esperada para a população de dias.

Figura 3.21: Exemplo de ajuste obtido via MLM, atendendo à hipótese que a relação $MP_{2.5} \times POA$ varia diariamente. Cores diferentes representam dados observados em dias diferentes.

Ajustamos o modelo (1.1) para cada ano do período do estudo utilizando a função `lmer()` do pacote `lme4` (Bates et al., 2015) do *software* R. Os comandos utilizados estão no Apêndice B. As estimativas dos parâmetros dos modelos, bem como seus erros-padrão, são apresentadas na seção 4.1.

3.3.2 Relação entre as concentrações de $MP_{2.5}$ e MP_{10} medidos na superfície e a POA do MODIS-MAIAC na RMSP

Ao contrário do que foi observado nos estudos que aplicaram a *Metodologia de Harvard* em outras regiões, a POA apresentou sinal negativo para a metade dos modelos ajustados na tese. Esses resultados são apresentados e discutidos na seção 4.1.

Na tentativa de encontrar possíveis erros no processamento dos dados de satélite e/ou dos dados da CETESB, o que poderia justificar os resultados inesperados encontrados, os códigos dos programas construídos para a compilação dos dados foram verificados, porém só foram encontradas algumas duplicações nos dados meteorológicos da CETESB.

A compilação dos dados do MODIS-MAIAC foi verificada com a ajuda do autor que validou esses dados para a América do Sul (Martins et al., 2017), que gentilmente compartilhou as imagens processadas por ele para comparação com as imagens processadas para esta tese. Nenhuma discrepância significativa foi encontrada.

Uma revisão de literatura mais detalhada mostrou que este estudo não foi o primeiro a encontrar relações fracas e/ou negativas entre a POA e o $MP_{2.5}$ e MP_{10} na RMSP. Natali (2008) estudou a relação entre a POA obtida por sensoriamento remoto e a concentração de $MP_{2.5}$ e MP_{10} na RMSP entre 2002 e 2005, e encontrou relações negativas entre eles. Natali (2008) utilizou o algoritmo desenvolvido por de Almeida Castanho (2005) que, com base nos dados de radiância do instrumento MODIS (Aqua e Terra, versão 4), tem como resultado

final dados de POA com uma resolução espacial de 1 km².

Utilizando os dados das redes de monitoramento automática (para MP₁₀) e manual (para MP_{2.5}) da CETESB, Natali (2008) estudou a influência dos seguintes fatores na relação do MP_{2.5} e do MP₁₀ com a POA: sazonalidade (estações seca e úmida), quantidade de água na coluna atmosférica, geometria do sensor (ângulo de espalhamento), dias com presença de nuvens e presença de aerossóis acima da CLP. Dentre esses fatores, o fator determinante foi a estrutura vertical da CLP.

Natali (2008) também utilizou os dados do sensor LIDAR (*Light Detection And Ranging*), que estima indiretamente a concentração relativa de aerossóis na atmosfera entre 300 m e 15 km, fornecendo informações sobre o perfil vertical da atmosfera. A conclusão foi que em dias sem a presença de aerossóis acima da CLP, observou-se um aumento nos valores da correlação entre a POA e o MP. Apesar do aumento observado nas correlações, essas permaneceram baixas (~ 0.30) e foram calculadas utilizando uma parte substancialmente menor da amostra inicial. Logo, tais conclusões precisam ser interpretadas com cautela.

Para tentar entender o resultado inesperado encontrado nesta tese, a relação entre o MP_{2.5} e o MP₁₀ com a POA (POA-MP) foi estudada de forma mais detalhada por meio da avaliação dos efeitos da sazonalidade (estações do ano), da ACLP e da umidade relativa (UR) na relação POA-MP. Além disso, utilizando dados da rede AERONET, comparamos algumas propriedades dos aerossóis de São Paulo e da costa leste dos Estados Unidos, com o intuito de verificar se o modelo de aerossol utilizado pelo algoritmo MAIAC (versão 6) na aquisição dos dados de POA sobre a RMSP é adequado.

Uma parte dos resultados dessa investigação foi apresentada na *2019 Air Pollution Conference Brazil - 4th CMAS South America* (Damascena et al., 2019). Outra parte foi publicada na Revista *Atmospheric Environment* em um artigo intitulado *Exploring the relationship between high-resolution aerosol optical depth values and ground-level particulate matter concentrations in the Metropolitan Area of São Paulo* (Anexo 1).

A seguir descrevemos a metodologia empregada nessa investigação.

3.3.2.1 Seleção das estações de monitoramento

Desde 2015, a CETESB classifica as estações de monitoramento de acordo com um critério chamado representatividade espacial. Essa classificação segue as recomendações do *Code of Federal Regulations - Title 40 - Protection of Environment/ Part 58 - Ambient Air Quality Surveillance/ Appendix E to Part 58 - Probe and Monitoring Path Siting Criteria for Ambient Air Quality Monitoring*, desenvolvido pela agência ambiental dos Estados Unidos (*United States Environmental Protection Agency* (USEPA)). A classificação considera o poluente, as características da região monitorada (e.g., topografia e características meteorológicas), bem como os impactos das emissões de fontes fixas e móveis (e.g, proximidade e intensidade) nas estações. Mais detalhes sobre essa classificação podem ser encontrados em CETESB (2016a).

Referente ao $MP_{2.5}$ e ao MP_{10} , como na RMSP a principal fonte dessas concentrações são as emissões veiculares, a distância entre a estação e a via de tráfego é uma das características consideradas na classificação por representatividade espacial. As estações são classificadas como micro, média, bairro e urbana. As duas primeiras são bastante impactadas por emissões locais enquanto as demais são representativas de áreas com dimensões variando entre $501 m^2$ e $50 km^2$ (CETESB, 2016a).

Nessa etapa da análise dos dados, foram utilizadas as concentrações de $MP_{2.5}$ e MP_{10} das estações de monitoramento classificadas como bairro e urbana, pois são as estações com representatividade espacial próxima ou maior do que a da POA obtida do MODIS-MAIAC ($1 km^2$). O número de estações que monitoraram o $MP_{2.5}$ e/ou o MP_{10} e que atenderam ao critério de representatividade espacial variou ao longo do período do estudo, como observado na Tabela 3.4.

Tabela 3.4: *Material particulado monitorado e representatividade espacial das estações de monitoramento no período do estudo*

Número da estação	Nome da estação	Representatividade espacial	2012	2013	2014	2015	2016	2017
1	Capão Redondo	bairro	MP_{10}	MP_{10}	MP_{10}	MP_{10}	MP_{10}	MP_{10}
2	Carapicuíba	bairro	MP_{10}	MP_{10}	MP_{10}	MP_{10}	MP_{10}	MP_{10}
3	USP	urbana	$MP_{2.5}$	$MP_{2.5}$	$MP_{2.5}$	$MP_{2.5}$	$MP_{2.5}$	$MP_{2.5}$
4	Diadema	bairro	MP_{10}	MP_{10}	MP_{10}	MP_{10}	MP_{10}	MP_{10}
5	Guarulhos - Paço Municipal	urbana	MP_{10}	MP_{10}	MP_{10}	MP_{10}	MP_{10}	MP_{10}
6	Guarulhos - Pimentas	bairro				MP	MP	MP
7	Ibirapuera	urbana			$MP_{2.5}$	$MP_{2.5}$	$MP_{2.5}$	$MP_{2.5}$
8	Interlagos	bairro	MP_{10}	MP_{10}	MP_{10}	MP_{10}	MP_{10}	MP_{10}
9	Itaim Paulista	urbana		MP_{10}	MP_{10}	MP	MP	MP
10	Mauá	bairro	MP_{10}	MP_{10}	MP_{10}	MP_{10}	MP_{10}	MP_{10}
11	Pq. D. Pedro II	bairro	MP_{10}	MP_{10}	MP_{10}	MP_{10}	MP	MP
12	Pico do Jaraguá	urbana					$MP_{2.5}$	$MP_{2.5}$
13	Santo André - Capuava	bairro	MP_{10}	MP_{10}	MP_{10}	MP_{10}	MP_{10}	MP_{10}
14	São Bernardo do Campo-Centro	bairro			$MP_{2.5}$	$MP_{2.5}$	$MP_{2.5}$	$MP_{2.5}$
15	São Bernardo do Campo-Paulicéia	bairro	MP_{10}	MP_{10}	MP_{10}	MP_{10}	MP_{10}	MP_{10}

MP: $MP_{2.5}$ e MP_{10}

3.3.2.2 Avaliação dos efeitos da ACLP, da UR e da sazonalidade na relação POA-MP

A relação POA-MP não é direta, pois é afetada por vários fatores, como a ACLP, UR e a distribuição vertical dos aerossóis (Chu et al., 2003; Li et al., 2015).

Como a POA é um valor associado à coluna atmosférica, a ACLP impacta a distribuição vertical dos aerossóis. Em geral, os aerossóis são bem misturados sob a CLP (Gupta et al., 2006; Koelemeijer et al., 2006; Guinot et al., 2006; Zhang et al., 2009a). No entanto, quando há o transporte de aerossol vindo de outras regiões, e em diferentes alturas, a relação POA-MP é afetada, pois as estações de monitoramento na superfície não capturam essas concentrações de MP em níveis mais altos. Além disso, as diferentes camadas de aerossóis na coluna atmosférica podem ter diferentes características físicas e diferentes composições

químicas, e como consequência, diferentes valores de POA (Slater and Dibb, 2004). Assim, a ACLP e a distribuição vertical dos aerossóis são variáveis-chave para a relação POA-MP.

Supondo que os aerossóis estejam bem misturados sob a CLP, estudos anteriores corrigiram os valores de POA obtidos de satélite utilizando a correção vertical descrita na equação (3.1) (Koelemeijer et al., 2006; Schaap et al., 2009; Boyouk et al., 2010; Chu et al., 2013; Tsai et al., 2011; Chew et al., 2016; Kong et al., 2016; Zheng et al., 2017; Gong et al., 2017). Essa correção geralmente melhora a correlação entre a POA e o MP, pois leva em consideração apenas o aerossol sob a CLP, eliminando o aerossol em níveis mais altos (Koelemeijer et al., 2006). Nesse caso, como não estamos considerando os aerossóis de toda a coluna atmosférica, o termo POA não é mais adequado e usaremos então o termo Aerossol Corrigido pela Altura (ACA).

$$ACA = \frac{POA}{ACLP} \quad (3.1)$$

As estimativas da ACLP no período da manhã e da tarde foram obtidas a partir de dados de sondagens atmosféricas, como descrito na seção 3.2.2.1.

A UR também é uma variável-chave na relação POA-MP. A CETESB utiliza o método de atenuação Beta para medir as concentrações de MP (MP_{10} e $MP_{2.5}$) (CETESB, 2019a; Xavier, 2011), no qual a quantificação do MP é determinada após o aquecimento do tubo de amostragem a uma temperatura acima da temperatura ambiente, com o intuito de remover a condensação e permitir a padronização de temperatura e da UR em todas as medições (Thermo Scientific, 2010). Portanto, medido dessa forma, o MP é uma “massa seca” de aerossol, enquanto a POA é medida em condições ambientais e é influenciada pela UR do ambiente. Dependendo da natureza higroscópica do aerossol, o aumento da UR pode modificar a composição química, a distribuição de tamanhos e as características de extinção do aerossol: as partículas dispersam mais luz, afetando os valores de POA (Liu et al., 2008).

A razão entre a dispersão do aerossol em um ambiente com uma UR específica e a dispersão do aerossol em uma condição de ar seco é denominado fator de crescimento higroscópico ($f(UR)$), que depende das características higroscópicas do aerossol, do período do dia e da região (Day and Malm, 2001). Então, uma outra correção, exibida na equação (3.2), minimiza o aumento dos valores da POA devido ao aumento da UR:

$$POA_{seca} = \frac{POA}{f(UR)} \quad (3.2)$$

em que POA_{seca} é o valor de POA corrigido pelo fator de crescimento higroscópico.

O fator de crescimento higroscópico é definido por meio de experimentos (Li et al., 2011). Na presente tese, utilizamos os valores de $f(UR)$ obtidos em experimentos realizados por Rodrigues (2014), entre 2013 e 2014, na Cidade de São Paulo. Considerando que $UR_0 = 40\%$, Rodrigues (2014) ajustou a função de crescimento higroscópico apresentada na equação (3.3)

(Hänel, 1976) aos dados gerados por técnicas de *Lidar Raman* (Ansmann et al., 1990).

$$f(UR) = \left(\frac{1 - \frac{UR}{100}}{1 - \frac{UR_0}{100}} \right)^{-\gamma} \quad (3.3)$$

A Tabela 3.5 apresenta as estimativas e os respectivos desvios-padrão para a taxa de crescimento do aerossol em condições de aumento da UR (γ) encontrados por Rodrigues (2014) em dias específicos do outono, primavera e verão.

Tabela 3.5: Taxa de crescimento do aerossol em condições de aumento da UR (γ) e respectivos desvios-padrão, de acordo com a estação do ano. Valores retirados do estudo de Rodrigues (2014).

estação do ano	γ	desvio-padrão
outono	0,11	0,03
primavera	0,48	0,12
verão	0,78	0,16

Para o período de estudo da tese, os valores de γ apresentados na Tabela 3.5 foram utilizados para corrigir os valores de POA obtidos no outono, primavera e verão, respectivamente. Por ser a estação do ano com características mais próximas às do inverno, utilizamos o mesmo valor de γ obtido no outono para corrigir os valores de POA obtidos no inverno.

Os valores de UR foram obtidos de 14 estações da CETESB e de 29 estações do Centro de Gerenciamento de Emergências Climáticas (CGE, <https://www.cgesp.org/v3/>). Para estações com valores de UR indisponíveis, foram atribuídos os valores de UR da estação mais próxima.

Foram calculados os coeficientes de correlação linear de *Pearson* entre as concentrações de $MP_{2.5}$ e MP_{10} com a POA do MODIS-MAIAC considerando as correções descritas nas equações 3.1 e 3.2, e também considerando a combinação de ambas as correções exibidas na equação (3.4).

$$ACA_{seca} = \frac{POA}{f(UR) \times ACLP} = \frac{POA_{seca}}{ACLP} \quad (3.4)$$

As correlações também foram calculadas de acordo com as estações do ano (sazonalidade), pois a relação POA-MP pode variar devido às condições meteorológicas (Boyounk et al., 2010).

Esses resultados são apresentados na seção 4.2.2.

3.3.2.3 Avaliação do modelo de aerossol utilizado pelo algoritmo MAIAC na aquisição de dados de POA sobre a RMSP

A obtenção dos dados de POA do MODIS é baseada na refletância da superfície e nas propriedades microfísicas das partículas do aerossol (Levy et al., 2007; de Almeida Castanho et al., 2008; Lyapustin et al., 2011b), que são representadas por diferentes modelos de aerossol

no algoritmo de aquisição de dados de POA. Como as propriedades microfísicas das partículas do aerossol mudam com as condições atmosféricas, quantidade e tipo de aerossol (Remer et al., 2005), a especificação correta dos modelos de aerossol no algoritmo é essencial para obter valores precisos de POA.

Dados de propriedades dos aerossóis (médias horárias) obtidos da rede AERONET (dados *Level 2*), no período de 2000 a 2011, foram utilizados para comparar propriedades dos aerossóis da costa leste dos EUA e de São Paulo. Essas duas estações foram escolhidas para a comparação, pois nas regiões onde estão localizadas, o algoritmo MAIAC utiliza o mesmo modelo de aerossol na aquisição da POA (Lyapustin et al., 2018). As características dos aerossóis da costa leste dos EUA são representadas pelos dados da estação GSFC (*NASA Goddard Space Flight Center*) que fica em *Greenbelt, Maryland* (Lyapustin et al., 2018). Já as características dos aerossóis de São Paulo são representadas pelos dados da estação São Paulo, localizada no topo do edifício mais alto do Instituto de Física que fica no campus da Universidade de São Paulo (Figura 3.1).

Os aerossóis de ambas as estações foram classificados utilizando as características albedo simples (AS) e fração da moda fina (FMF) como proposto por Lee et al. (2010). O AS define a capacidade relativa do aerossol absorver radiação em um determinado comprimento de onda: quanto maior o AS, menor a absorção de radiação (Lee et al., 2010). Já a FMF define a proporção de partículas finas (partículas com diâmetro menor que $1 \mu m$) do aerossol. De acordo com a classificação proposta por Lee et al. (2010), aerossóis com FMF maior que 0,6 são aerossóis da moda fina e aerossóis com FMF menor que 0,4 são aerossóis da moda grossa. Aerossóis com FMF entre 0,4 e 0,6 são classificados como uma mistura de aerossóis das modas grossa e fina.

Dependendo da natureza higroscópica do aerossol, a UR pode modificar suas características de absorção e espalhamento, e conseqüentemente alterar os valores de POA. Dessa forma, utilizamos a característica Total de Água Precipitável (TAP) como uma substituta para a UR na coluna atmosférica para avaliar como a UR altera o AS do aerossol em São Paulo e na costa leste dos Estados Unidos.

Esses resultados são apresentados na seção 4.2.3.

3.3.3 Modelo proposto para a predição das concentrações de $MP_{2.5}$ na RMSP

Para o problema que estamos estudando na presente tese, a principal contribuição da metodologia desenvolvida pelos pesquisadores de Harvard foi a proposta de utilizar o MLM e os dados de POA na predição das concentrações de $MP_{2.5}$. Porém, como será demonstrado nas seções 4.2.2 e 4.2.3, não recomendamos que os dados de POA da versão 6 do MODIS-MAIAC sejam utilizados como uma aproximação para as concentrações de $MP_{2.5}$ na superfície da RMSP. No entanto, o MLM ainda pode ser utilizado para estimar as concentrações diurna e diária de $MP_{2.5}$ em cada ponto da grade esquematizada na Figura 3.2 no período do estudo.

Nessa seção, propomos uma metodologia baseada apenas nas etapas 1 e 2 da *Metodologia de Harvard*. Pensando no contexto da RMSP, no modelo do primeiro estágio da metodologia proposta, além das variáveis meteorológicas e de uso e ocupação do solo, incluímos uma variável indicadora da ocorrência de feriado, variáveis indicadoras dos dias da semana e termos específicos para capturar a sazonalidade das concentrações de $MP_{2.5}$ ao longo do tempo. Além disso, propomos modelos que consideram as duas fontes de correlação, espacial e temporal, como introduzido na seção 1.7.

A metodologia proposta é detalhada a seguir.

3.3.3.1 Primeiro estágio

Além de não utilizarmos os dados de POA no primeiro estágio, também não utilizamos a variável ACLP. Como descrito na seção 3.2.2.2.1, os dados de ACLP obtidos do modelo do NCEP apresentaram uma baixa concordância com os dados de ACLP obtidos de sondagem atmosférica. Já os dados de sondagem atmosférica apresentaram muitos dados faltantes, o que diminui o número de observações para o ajuste do modelo.

Considerando essas alterações, a equação (1.1) foi reescrita como:

$$MP_{it} = \alpha + u_t + \beta_1 Temperatura_{it} + \beta_2 UmidadeRelativa_{it} + \beta_3 Precipitacao_{it} + \beta_4 VelocidadeVento_{it} + \beta_5 DensidadeViaria_i + \epsilon_{it} \quad (3.5)$$

em que MP_{it} corresponde à média diária da concentração de $MP_{2.5}$ no dia t na célula i ; α é o intercepto fixo; u_t é o intercepto aleatório de dia; os β_s são os parâmetros associados às variáveis da parte fixa do modelo; $Temperatura_{it}$, $UmidadeRelativa_{it}$, $Precipitacao_{it}$ e $VelocidadeVento_{it}$ são os valores de temperatura da superfície, umidade relativa, precipitação acumulada e velocidade do vento correspondentes à célula i no dia t ; $DensidadeViaria_i$ é o valor da densidade viária observado na célula i ; supõe-se $\epsilon_{it} \sim N(0, \sigma^2)$ e $u_t \sim N(0, \sigma_u^2)$, mutuamente independentes.

Para efeito de comparação, considere os conjuntos de variáveis explicativas apresentadas no Tabela 3.6. Assim como na *Metodologia de Harvard*, o conjunto de variáveis explicativas E1 considera apenas as variáveis meteorológicas e de uso e ocupação do solo. O conjunto E2, além das variáveis do conjunto E1, considera a variável indicadora de ocorrência de feriado (Feriado), as variáveis indicadoras dos dias da semana (Dia da Semana), bem como os termos referentes às interações entre os dias da semana e a ocorrência de feriado (Feriado:Dia da Semana). O conjunto E3, além das variáveis do conjunto E2, considera termos senos e cossenos com períodos que são funções do tempo (dias) para capturar a sazonalidade das concentrações de $MP_{2.5}$ ao longo do tempo.

Tabela 3.6: Conjuntos de variáveis explicativas utilizadas nos modelos ajustados na tese.

E1	E2	E3
Temperatura	Temperatura	Temperatura
Umidade	Umidade	Umidade
Velocidade do Vento	Velocidade do Vento	Velocidade do Vento
Precipitação	Precipitação	Precipitação
Densidade Viária	Densidade Viária	Densidade Viária
	Feriado	Feriado
	Dia da Semana	Dia da Semana
	Feriado:Dia da Semana	Feriado:Dia da Semana
	Seno	
	Cosseno	

Como pode ser observado na Figura 3.22, a média das concentrações diárias de $MP_{2.5}$, de uma forma geral, diminui nos feriados, provavelmente devido ao número de veículos que deixam de circular na cidade. Essa diminuição pode depender do dia da semana em que os feriados ocorrem: feriados na segunda/terça ou na quinta/sexta motivam as chamadas "emendas de feriado" e com isso a diminuição nas concentrações deve ser maior do que a observada quando o feriado ocorre em uma quarta-feira. Os efeitos de interação da variável indicadora da ocorrência de feriado com as variáveis indicadoras de dia da semana servem para capturar esse comportamento.

Figura 3.22: Média das concentrações diárias de $MP_{2.5}$ observadas nas estações de monitoramento automática da CETESB em 2017. A linha vermelha representa a suavização da tendência das concentrações. As linhas verticais tracejadas cinza indicam os dias em que ocorreram feriados no ano de 2017.

O padrão sazonal da média das concentrações diárias de $MP_{2.5}$ também pode ser observado na Figura 3.22. Note o aumento das concentrações nos meses de inverno e primavera. Para capturar esse comportamento incluímos nos modelos as funções periódicas seno e cosseno com períodos que são funções do tempo (dias). O número de termos incluídos é proporcional ao número de ciclos observados (Morettin, 2017).

Os conjuntos de variáveis E1, E2 e E3 foram organizados dessa forma, pois temos interesse em verificar as seguintes hipóteses:

1. por estarem relacionadas à dinâmica das emissões veiculares na RMSP, acreditamos que as variáveis feriado e as indicadoras dos dias da semana, bem como a interação entre elas, além das variáveis meteorológicas e de uso e ocupação do solo (conjunto E2), são variáveis que podem melhorar o poder de predição do $MP_{2.5}$ quando comparado ao conjunto de variáveis E1;
2. além das variáveis do conjunto E2, os termos específicos para capturar a sazonalidade da concentração de $MP_{2.5}$ (senos e cossenos), são variáveis que podem melhorar o poder de predição do $MP_{2.5}$ quando comparado ao conjunto de variáveis E1.

Além disso, no problema que estamos estudando, existem dois tipos de correlação que devem consideradas na estimação dos parâmetros do modelo: a espacial e a temporal. A inclusão dos efeitos aleatórios u_t (equação (3.5)) contempla a correlação existente entre as observações de um mesmo dia (correlação espacial). No entanto, o modelo (3.5) ignora a correlação existente entre as observações de uma mesma estação de monitoramento ao longo do tempo (correlação temporal). Tal correlação é considerada na metodologia proposta na tese ao incluirmos na equação (3.5) os efeitos aleatórios de estação, s_i , como exemplificado na equação (3.6):

$$MP_{it} = \alpha + u_t + s_i + \beta_1 Temperatura_{it} + \beta_2 UmidadeRelativa_{it} + \beta_3 Precipitacao_{it} + \beta_4 VelocidadedoVento_{it} + \beta_5 DensidadeViaria_i + \epsilon_{it} \quad (3.6)$$

em que MP_{it} corresponde à média diária da concentração de $MP_{2.5}$ no dia t na célula i ; α é o intercepto fixo; u_t é o intercepto aleatório de dia; s_i é o intercepto aleatório de estação de monitoramento; os β_s são os parâmetros associados às variáveis da parte fixa do modelo; $Temperatura_{it}$, $UmidadeRelativa_{it}$, $Precipitacao_{it}$ e $VelocidadedoVento_{it}$ são os valores de temperatura da superfície, umidade relativa, precipitação acumulada e velocidade do vento correspondentes à célula i no dia t ; $DensidadeViaria_i$ é o valor da densidade viária observado na célula i ; supõe-se $\epsilon_{it} \sim N(0, \sigma^2)$, $u_t \sim N(0, \sigma_u^2)$ e $s_i \sim N(0, \sigma_s^2)$, mutuamente independentes.

Para cada conjunto de variáveis explicativas apresentadas na Tabela 3.6, ajustamos MLMs que consideram apenas a fonte de correlação espacial, aos quais vamos nos referir como modelos do tipo A1 e modelos que consideram as correlações espacial e temporal, aos quais vamos nos referir como modelos do tipo A2. Dessa forma, queremos verificar a terceira hipótese:

3. os modelos do tipo A2 apresentam maior poder de predição quando comparados aos modelos do tipo A1, pois consideram as duas fontes de correlação existente nos dados: a espacial, por meio da inclusão dos efeitos aleatórios u_t , e a temporal, por meio da inclusão dos efeitos aleatórios s_i .

A esquematização e a definição das equações dos modelos ajustados são apresentados no Tabelas 3.7 e C.1 (Apêndice C), respectivamente.

Tabela 3.7: *Variáveis explicativas e efeitos aleatórios incluídos nos modelos ajustados*

	Modelo					
	E1A1	E2A1	E3A1	E1A2	E2A2	E3A2
Temperatura	X	X	X	X	X	X
Umidade	X	X	X	X	X	X
Velocidade do Vento	X	X	X	X	X	X
Precipitação	X	X	X	X	X	X
Densidade Viária	X	X	X	X	X	X
Feriado		X	X		X	X
Dia da Semana		X	X		X	X
Feriado:Dia da Semana		X	X		X	X
Senos			X			X
Cossenos			X			X
u_t	X	X	X	X	X	X
s_i				X	X	X

Assim como [Just et al. \(2015\)](#), que aplicaram a *Metodologia de Harvard* para dados da Cidade do México, utilizamos a série completa de dados (2012-2017) nos ajustes dos MLMs ao invés de considerar os ajustes ano a ano, como proposto na *Metodologia de Harvard*.

Para ajustar os modelos esquematizados na Tabela 3.7 utilizamos a função `lmer()` do pacote `lme4` ([Bates et al., 2015](#)) do *software* R. Os *scripts* utilizados no ajuste dos modelos do tipo A1 e A2 estão nos Apêndices D e E, respectivamente.

3.3.3.1.1 Seleção de variáveis nos modelos ajustados no primeiro estágio

Para cada um dos seis modelos, as variáveis incluídas nos respectivos modelos finais foram selecionadas como sugerido por [Pinheiro and Bates \(2000\)](#) utilizando a função `step()` do pacote `lmerTest` ([Kuznetsova et al., 2017](#)). Fixadas as variáveis explicativas do modelo selecionou-se a estrutura de variâncias e covariâncias por meio do Teste da Razão Verossimilhanças (TRV) utilizando a função de verossimilhança restrita. Selecionada a estrutura de variâncias e covariâncias, as variáveis explicativas foram selecionadas por meio do TRV utilizando a função de verossimilhança.

O nível de significância considerado nos testes foi o de 5%.

3.3.3.1.2 Comparação dos modelos ajustados no primeiro estágio

Os modelos finais, obtidos após o ajuste e seleção das variáveis para cada um dos modelos esquematizados na Tabela 3.7, foram comparados por meio do Critério de Informação de Akaike (AIC, [Akaike \(1974\)](#)), muito utilizado na comparação de modelos encaixados ou não ¹. O AIC é um critério que penaliza a verossimilhança com o número de parâmetros

¹Dizemos que um modelo M1 é encaixado no modelo M2 quando M1 é um caso particular ou uma redução de M2.

estimados e é uma medida de desempenho preditivo de um modelo estatístico. Os modelos mais adequados são aqueles com menores valores de AIC.

No contexto dos MLMs, temos o AIC marginal e o AIC condicional. O primeiro, que é o automaticamente calculado pelos pacotes computacionais que ajustam MLMs, não leva em conta a predição dos efeitos aleatórios, enquanto que o segundo leva. Qual utilizar depende de qual tipo de predição é de interesse. Quando o interesse está em avaliar o desempenho preditivo do modelo para as unidades amostrais do próprio estudo usa-se o AIC condicional, pois para essas unidades conhecemos as predições dos efeitos aleatórios. Se o interesse está em avaliar o desempenho preditivo do modelo para as novas unidades amostrais o AIC marginal é o mais adequado, pois para novas unidades amostrais não é possível prever os efeitos aleatórios (Vaida and Blanchard, 2005).

No contexto deste estudo, o interesse está em realizar predições para células sem estações de monitoramento em instantes t já observados. Nessas condições, células sem estação de monitoramento são novas unidades amostrais, para as quais não é possível prever os efeitos aleatórios s_i , mas para as quais podemos prever os efeitos aleatórios u_t , uma vez que os instantes t já foram observados. Dessa forma, para os modelos do tipo A1 e A2 esquematizados na Tabela 3.7, as predições pontuais de $MP_{2.5}$ para as células sem estações de monitoramento nos instantes t levaram em conta apenas os efeitos aleatórios u_t .

De acordo com Vaida and Blanchard (2005), por levar em conta as predições dos efeitos aleatórios u_t , o ideal seria utilizarmos o AIC condicional na comparação dos modelos ajustados no primeiro estágio. Porém, devido às limitações computacionais, não conseguimos calculá-lo. Como nossa série de dados é relativamente grande (~ 6 mil observações) e as operações envolvidas no cálculo do AIC condicional demandam um alto desempenho computacional, mesmo utilizando computadores com 8, 16 e 32 GB de memória RAM, não tivemos êxito. Sendo assim, os modelos foram comparados por meio do AIC marginal.

3.3.3.1.3 Validação e avaliação do desempenho dos modelos ajustados no primeiro estágio

Os modelos finais obtidos a partir dos modelos esquematizados na Tabela 3.7 foram validados por meio do Critério de Validação Cruzada. Para cada modelo, esse procedimento consiste em: os dados de uma determinada estação de monitoramento são excluídos do conjunto de dados e um novo modelo é ajustado considerando os dados das demais estações; com as estimativas obtidas desse novo modelo faz-se a predição dos valores de $MP_{2.5}$ para aquela estação que foi retirada da amostra; repete-se esse procedimento até que todas as estações tenham seu dados preditos.

Ao final desse processo foi calculado o valor do coeficiente de determinação (R^2) da regressão dos valores observados de $MP_{2.5}$ contra os valores preditos na etapa de Validação Cruzada. O R^2 mostra o quanto da variabilidade do $MP_{2.5}$ foi explicada pelos modelos de calibração na etapa de Validação Cruzada e varia entre 0 e 1. Valores próximos de 1 indicam

que o modelo consegue explicar grande parte da variabilidade do $MP_{2.5}$.

O desempenho dos modelos finais derivados dos modelos esquematizados na Tabela 3.7 também foi avaliado por meio do cálculo da raiz quadrada do erro quadrático médio de predição (REQMP), apresentado na equação (3.7):

$$REQMP = \sqrt{\frac{\sum_{i=1}^N (MP_{observado} - MP_{predito})^2}{N}} \quad (3.7)$$

em que N é o número total de observações. Tanto o R^2 quanto o $REQMP$ foram calculados considerando os dados de todas as estações de monitoramento, quanto os dados de cada estação individualmente.

Esses resultados são apresentados na seção 4.3.1.

3.3.3.2 Segundo estágio

Nessa etapa, os modelos finais obtidos a partir dos modelos esquematizados na Tabela 3.7 foram utilizados para prever as concentrações de $MP_{2.5}$ nas células da grade esquematizada na Figura 3.2 que não possuem estações de monitoramento no período do estudo.

Além das predições pontuais, precisamos apresentar uma medida de incerteza associada a essas predições ², o que envolve o cálculo do desvio-padrão associado ao erro de predição. Até o fechamento da presente tese, não havia disponível no *software* R um pacote que o calculasse de forma que atendesse às necessidades da presente tese. Para garantir que tanto as predições pontuais quanto os desvios-padrão associados aos erros de predição fossem calculados corretamente, desenvolvemos uma rotina computacional para a realização dos cálculos. O cálculo dos desvios-padrão associados aos erros de predição, no *Modelo de Harvard* e no modelo proposto, é uma das contribuições da presente tese e está detalhado no Apêndice F.

Após os cálculos das predições pontuais, foram construídos mapas com as médias anual e sazonais das concentrações de $MP_{2.5}$ preditas pelos modelos finais. De modo análogo, apresentamos os mapas com a média dos desvios-padrão associados aos erros de predição no período do estudo. Esses resultados são apresentados na seção 4.3.2.

Utilizamos o *software* R nessas análises. Os *scripts* utilizados no ajuste dos modelos, na etapa de validação cruzada, na predição das concentrações de $MP_{2.5}$, no cálculo dos desvios-padrão dos erros de predição e na confecção dos mapas de concentração, para os modelos do tipo A1 e A2, estão nos Apêndices D e E, respectivamente.

Resultados parciais referentes aos modelos propostos foram apresentados na versão *online* da *12th International Conference on Air Quality Science and Application* (Damascena et al., 2020).

²“The Statistician is the Wizard who makes “scientific” statements about invisible states and quantities. However, contrary to the real wishes (or witches), (s)he attaches uncertainties to (her)his statements.” Carlos Alberto de Bragança Pereira

3.4 Questões Éticas

Apesar de utilizar apenas dados secundários com acesso público, o presente projeto de pesquisa foi aprovado pela Comissão de Ética em Pesquisa da Faculdade de Medicina da Universidade de São Paulo, em 06/04/2016 (Anexo [2](#)).

4. Resultados

4.1 Primeiro estágio da *Metodologia de Harvard* para a predição da concentração de $MP_{2.5}$ na RMSP

Para cada ano do período do estudo, as Tabelas 4.1 e 4.2 apresentam, respectivamente, as estimativas dos parâmetros do modelo (1.1), bem como seus erros-padrão, quando considerados os dados dos satélites Terra e Aqua e as concentrações diárias de $MP_{2.5}$.

Observamos que para a metade dos modelos ajustados os sinais associados à POA do MODIS-MAIAC foram negativos. Os resultados dos estudos que aplicaram essa metodologia em outras regiões do mundo apresentaram em seus modelos sinais positivos para a POA (Estados Unidos: Lee et al. (2011), Kloog et al. (2011), Kloog et al. (2012) e Chudnovsky et al. (2014b); Europa: (Stafoggia et al., 2017; de Hoogh et al., 2018) ; China: Xiao et al. (2017) ; México: Just et al. (2015) ; Israel: Kloog et al. (2015)).

Tabela 4.1: *Estimativas (erros-padrão) obtidos para o modelo descrito na equação (1.1) considerando os dados do satélite Terra e as médias das concentrações de $MP_{2.5}$.*

	2012	2013	2014	2015	2016	2017
Intercepto	32,603 (2,932)	21,254 (4,572)	13,815 (4,190)	30,927 (3,943)	31,865 (3,110)	27,678 (2,011)
POA TERRA	4,140 (3,230)	-16,814 (3,748)	0,643 (3,201)	4,092 (3,156)	-1,146 (2,576)	-7,219 (3,241)
Densidade Viária	2,682 (0,480)	0,259 (0,573)	0,668 (0,509)	0,816 (0,537)	4,647 (0,512)	6,377 (0,498)
ACLP TERRA	0,002 (0,001)	-0,002 (0,001)	0,002 (0,001)	0,001 (0,001)	0,000 (0,001)	0,003 (0,001)
Precipitação acum	-0,002 (0,012)	0,019 (0,022)	-0,010 (0,017)	-0,022 (0,016)	0,019 (0,024)	-0,029 (0,021)
Temperatura	-0,184 (0,083)	0,219 (0,140)	0,191 (0,134)	-0,361 (0,126)	-0,569 (0,102)	-0,260 (0,062)
Umidade	-0,148 (0,020)	-0,010 (0,028)	0,022 (0,025)	-0,054 (0,022)	-0,042 (0,017)	-0,073 (0,013)
Velocidade Vento	0,469 (0,193)	-0,293 (0,203)	-0,136 (0,190)	-1,378 (0,210)	-0,171 (0,159)	-0,716 (0,148)

Tabela 4.2: *Estimativas (erros-padrão) obtidas para o modelo descrito na equação (1.1) considerando os dados do satélite Aqua e as médias diárias das concentrações de $MP_{2.5}$.*

	2012	2013	2014	2015	2016	2017
Intercepto	34,281 (4,063)	20,881 (5,436)	10,091 (5,218)	31,071 (5,122)	20,429 (3,487)	23,146 (2,514)
POA AQUA	6,418 (3,805)	-15,401 (4,401)	1,021 (3,153)	2,121 (4,267)	-5,144 (3,143)	-0,783 (3,268)
Densidade Viária	0,835 (0,717)	1,078 (0,752)	1,065 (0,671)	1,364 (0,829)	5,493 (0,691)	6,917 (0,637)
ACLP AQUA	0,002 (0,001)	0,000 (0,001)	0,003 (0,001)	0,002 (0,001)	0,004 (0,001)	0,004 (0,001)
Precipitacao acum	-0,084 (0,049)	-0,053 (0,047)	-0,064 (0,065)	-0,024 (0,039)	-0,141 (0,063)	-0,048 (0,029)
Temperatura	-0,202 (0,116)	0,320 (0,168)	0,106 (0,166)	-0,361 (0,160)	-0,383 (0,118)	-0,333 (0,077)
Umidade	-0,174 (0,030)	-0,040 (0,035)	0,077 (0,032)	-0,061 (0,030)	0,001 (0,022)	-0,048 (0,018)
Velocidade Vento	0,008 (0,235)	-0,738 (0,253)	-0,130 (0,223)	-1,580 (0,295)	0,033 (0,183)	-0,818 (0,170)

4.2 Relação entre as concentrações de $MP_{2.5}$ e MP_{10} medidas na superfície e a POA obtida do MODIS-MAIAC na RMSP

Nessa etapa do estudo, devido aos sinais negativos associados à POA nos modelos ajustados na seção 4.1, a relação linear entre a POA e as concentrações horária, diurna e diária de $MP_{2.5}$ e MP_{10} medidos na superfície foi estudada de forma mais aprofundada.

4.2.1 Correlação entre as concentrações de $MP_{2.5}$ e MP_{10} na superfície e os valores de POA do MODIS-MAIAC na RMSP

A Tabela 4.3 apresenta os valores de correlação entre as concentrações horária, diurna e diária de MP_{10} e $MP_{2.5}$ medidos na superfície e os valores de POA do MODIS-MAIAC, no período completo do estudo e nas estações do ano, considerando os dados das 15 estações de monitoramento utilizadas nessa etapa do estudo (seção 3.3.2). No geral, as correlações foram negativas e/ou baixas tanto para o $MP_{2.5}$ quanto para o MP_{10} . Os maiores valores de correlação foram observados entre os dados de POA e as concentrações de MP no momento de passagem do satélite Aqua. As correlações entre a POA no momento de passagem do satélite Terra e as concentrações de MP (horária, diurna e diária) foram majoritariamente negativas. Referente às estações do ano, a primavera (SON) foi a que apresentou as maiores correlações.

As diferenças observadas entre as correlações dos dados dos dois satélites podem ser explicadas pela mistura de aerossóis na coluna atmosférica. O Terra passa sobre a RMSP durante o meio da manhã, quando a CLP começa a evoluir e, como consequência, os aerossóis ainda não estão bem misturados na coluna atmosférica. Em contraste, o Aqua passa sobre a RMSP no início da tarde, quando a CLP já está desenvolvida e os aerossóis já estão bem misturados na coluna atmosférica.

Além disso, alguns estudos demonstraram que a qualidade dos dados do MODIS a bordo do Terra piorou na coleção 5, devido à degradação da banda azul do sensor (Levy et al., 2010; Wang et al., 2012; DeVisser and Messina, 2013; Zhang et al., 2017). Mesmo após as melhorias realizadas na coleção 6 em 2014, a degradação desse sensor afeta os valores de POA do MAIAC “por meio de listras e um viés positivo no lado esquerdo da varredura, principalmente sobre superfícies brilhantes” (Lyapustin et al., 2018).

Tabela 4.3: Correlação linear (R) entre as concentrações horária, diurna e diária de MP_{10} e $MP_{2.5}$ medidos na superfície e os valores de POA do MODIS-MAIAC, no período do estudo (2012-2017) e de acordo com as estações do ano.

Intervalo	Satélite	Período/estação	MP_{10}		$MP_{2.5}$	
			n	R	n	R
Horário	Aqua	2012-2017	3563	0,201	1345	0,229
		DJF	230	0,236	50	0,284
		MAM	503	0,243	174	0,144
		JJA	1949	0,179	759	0,241
		SON	881	0,249	362	0,245
	Terra	2012-2017	5681	-0,031	1942	0,021
		DJF	1019	-0,119	268	-0,03
		MAM	1159	-0,03	352	0,003
		JJA	2086	-0,089	768	-0,056
		SON	1417	0,111	554	0,12
Diurno	Aqua	2012-2017	3547	0,088	1324	0,157
		DJF	228	0,001	46	0,12
		MAM	501	0,157	170	0,044
		JJA	1944	0,076	758	0,156
		SON	874	0,192	350	0,219
	Terra	2012-2017	5755	0,077	1933	0,098
		DJF	1036	-0,058	253	0,022
		MAM	1176	0,074	360	0,065
		JJA	2115	-0,011	771	0,084
		SON	1428	0,241	549	0,17
Diário	Aqua	2012-2017	3503	-0,068	1299	-0,013
		DJF	223	-0,09	47	0,074
		MAM	500	0,022	165	-0,052
		JJA	1915	-0,052	744	0,015
		SON	865	0,082	343	0,08
	Terra	2012-2017	5700	-0,035	1915	-0,013
		DJF	1036	-0,097	249	0,026
		MAM	1169	-0,016	365	-0,045
		JJA	2083	-0,111	763	0,008
		SON	1412	0,157	538	0,051

DJF: Dezembro, Janeiro, Fevereiro; MAM: Março, Abril, Maio; JJA: Junho, Julho, Agosto; SON: Setembro, Outubro, Novembro.

4.2.2 Avaliação dos efeitos da ACLP, da UR e da sazonalidade na relação POA-MP

As correlações entre as concentrações horária, diurna e diária de $MP_{2.5}$ e MP_{10} medidos na superfície e a POA do MODIS-MAIAC corrigida como descrito nas equações (3.1), (3.2) e (3.4) são apresentadas na Tabela 4.4. Devido aos dados faltantes relacionados às estimativas

da ACLP (seção 3.2.2.2.1), as correlações da Tabela 4.3 foram recalculadas e apresentadas na Tabela 4.4 para comparação.

Após o recálculo das correlações apresentadas na Tabela 4.3 o padrão dos resultados não mudou. Aumentos discretos nos valores das correlações foram observados após a correção pela ACLP (equação (3.1)), principalmente para os dados de POA do satélite Terra. Também foram observados aumentos nos valores das correlações após a correção pelo fator de crescimento higroscópico ($f(UR)$) (equação (3.2)), principalmente para os dados de POA do satélite Aqua. Os aumentos observados nos valores das correlações após a combinação de ambas as correções (equação (3.4)), em geral, foram maiores dos que os observados com apenas uma das correções. Apesar dos aumentos, os valores de correlação permaneceram abaixo de 0,40, para ambos os satélites e para todas as concentrações de $MP_{2.5}$ e MP_{10} (horária, diurna e diária).

Os resultados relacionados às correções nos dados de POA do satélite Aqua devem ser interpretados com cautela, pois mais de 50% dos dados eram faltantes após a aplicação das correções apresentadas nas equações (3.1), (3.2) e (3.4), devido à correção pela ACLP.

Para ambos os satélites e MPs, as correlações sazonais foram tipicamente mais fortes do que as correlações anuais, mas ainda fracas, mesmo após as correções pela ACLP e pelo fator $f(UR)$.

Tabela 4.4: Correlação linear (R) entre as concentrações horária, diurna e diária de MP_{10} e $MP_{2.5}$ medidos na superfície e os valores de POA do MODIS-MAIAC, corrigidos de acordo com as equações (3.1), (3.2) e (3.4), para ambos os satélites e de acordo com as estações do ano.

PM	Intervalo	Período/estação	Aqua				Terra					
			n	POA	ACA	POA _{seca}	ACA _{seca}	n	POA	ACA	POA _{seca}	ACA _{seca}
MP ₁₀	horário	2012-2017	1649	0,176	0,195	0,265	0,28	3480	-0,076	0,153	-0,007	-0,02
		DJF	103	0,319	0,317	0,346	0,277	623	-0,184	-0,142	-0,033	-0,021
		MAM	231	0,32	0,298	0,338	0,314	669	-0,032	0,188	0,001	0,234
		JJA	850	0,11	0,122	0,146	0,168	1087	-0,074	0,16	-0,025	0,171
		SON	465	0,214	0,248	0,328	0,353	1101	0,014	0,176	0,033	0,099
	diurno	2012-2017	1637	0,01	0,009	0,235	0,234	3522	0,037	0,226	0,177	0,183
		DJF	101	-0,029	0,197	0,339	0,381	633	-0,115	-0,133	0,097	0,037
		MAM	231	0,244	0,222	0,278	0,255	678	0,054	0,254	0,086	0,276
		JJA	846	-0,036	-0,108	0	-0,058	1105	0,039	0,23	0,078	0,251
		SON	459	0,125	0,127	0,29	0,307	1106	0,152	0,289	0,27	0,332
	diário	2012-2017	1632	-0,145	-0,16	0,15	0,129	3505	-0,077	0,216	0,114	0,145
		DJF	100	-0,183	0,071	0,291	0,306	634	-0,129	-0,148	0,077	0,041
		MAM	232	0,086	0,072	0,13	0,112	675	-0,047	0,235	-0,015	0,259
		JJA	842	-0,157	-0,264	-0,124	-0,218	1095	-0,1	0,207	-0,058	0,235
		SON	458	0,031	0,011	0,223	0,226	1101	0,076	0,275	0,246	0,352
MP _{2.5}	horário	2012-2017	549	0,241	0,29	0,264	0,318	1014	0,034	0,064	0,236	0,025
		DJF	17	0,272	0,142	0,193	0,017	124	-0,093	0,161	-0,1	0,163
		MAM	43	0,267	0,275	0,292	0,295	117	0,073	0,159	0,318	0,399
		JJA	269	0,176	0,208	0,181	0,223	333	-0,005	0,027	0,225	0,196
		SON	220	0,321	0,355	0,335	0,385	440	0,084	-0,019	0,23	-0,011
	diurno	2012-2017	535	0,081	0,232	0,035	0,192	1014	0,13	0,259	0,345	0,323
		DJF	18	0,126	0,016	0,031	-0,116	115	0,064	0,233	-0,001	0,177
		MAM	42	0,074	0,094	0,198	0,215	121	0,01	0,046	0,29	0,333
		JJA	269	-0,033	0,004	-0,104	-0,051	337	0,155	0,192	0,334	0,346
		SON	206	0,233	0,359	0,139	0,323	441	0,164	0,278	0,346	0,373
	diário	2012-2017	510	-0,074	0,154	-0,168	0,057	983	-0,017	0,181	0,307	0,295
		DJF	18	0,287	0,087	0,243	0,021	106	0,126	0,272	0,032	0,205
		MAM	36	0,033	0,059	-0,009	0,016	116	-0,086	-0,041	0,217	0,267
		JJA	255	-0,181	-0,144	-0,278	-0,223	331	0,007	0,056	0,325	0,339
		SON	201	0,089	0,276	-0,079	0,191	430	0,026	0,201	0,272	0,359

DJF: Dezembro, Janeiro, Fevereiro; MAM: Março, Abril, Maio; JJA: Junho, Julho, Agosto; SON: Setembro, Outubro, Novembro

4.2.3 Avaliação do modelo de aerossol utilizado pelo algoritmo MAIAC na aquisição de dados de POA sobre a RMSP

Na Figura 4.1 apresentamos a classificação dos aerossóis das estações GSFC (que representa as características dos aerossóis da costa leste dos EUA) e São Paulo, como proposta por Lee et al. (2010), e que foi detalhada na seção 3.3.2.3.

As colunas atmosféricas de São Paulo e da costa leste dos EUA são dominadas por aerossóis da moda fina. Porém, os aerossóis na costa leste dos EUA apresentam valores de albedo simples (AS) relativamente homogêneos, enquanto os aerossóis em São Paulo apresentam alta variabilidade nos valores de AS.

Apesar das diferenças entre essas propriedades dos aerossóis em ambas as localidades, o modelo de aerossol especificado para a América do Sul pelo algoritmo MAIAC é o mesmo modelo especificado para a costa leste dos EUA (ver Figura 4 em Lyapustin et al. (2018)), demonstrando que existe uma discordância entre o modelo de aerossol especificado pelo algoritmo MAIAC e as propriedades dos aerossóis na RMSP.

Figura 4.1: Classificação proposta por Lee et al. (2010) mostrando a dispersão entre o AS e a FMF, de acordo com as estações do ano, para as estações GSFC (esquerda) e São Paulo (direita) da rede AERONET, no período de 2000-2011. FA-A: fino altamente absorvente, FM-A: fino moderadamente-absorvente, FL-A: fino levemente-absorvente, FN-A: fino não-absorvente, GN-A: grosso não-absorvente, M-A: misto absorvente, MN-A: misto não-absorvente.

O efeito da UR na relação POA-MP também pode estar relacionado à especificação dos modelos de aerossol do MAIAC, já que a UR pode alterar as características químicas e de extinção dos aerossóis (comunicação via e-mail com Alexei Lyapustin realizada em 2019).

Como os sulfatos e nitratos, que são altamente higroscópicos, são os principais constituintes do aerossol em São Paulo (Martins Pereira et al., 2017) e no leste dos EUA (Bell et al., 2007), é possível que o aumento da UR altere o AS do aerossol na RMSP, tornando-

o comparável ao do aerossol no leste dos EUA. Isso explicaria porque o MAIAC utiliza o mesmo modelo de aerossol nos dois locais, apesar das diferenças no AS observadas na Figura 4.1.

Para investigar essa hipótese, construímos a Figura 4.2 que apresenta a média do AS para cada *bin* de FMF e Total de Água Precipitável (TAP) nas estações GSFC e São Paulo. As mudanças no TAP são causadas em parte pela UR e em parte pela variação sazonal, já que o TAP é uma função da temperatura e da UR (Lee et al., 2010). O aumento causado pelo TAP no AS em São Paulo é menor do que o aumento causado pelo TAP no AS na costa leste dos EUA. Portanto, mesmo levando em consideração o efeito da UR, o aerossol em São Paulo é mais absorvente do que o aerossol da costa leste dos EUA, evidenciando que o modelo de aerossol aplicado à RMSP pela versão 6 do algoritmo MAIAC não está adequado.

Sulfatos e nitratos são altamente refletores de radiação solar (Wallace and Hobbs, 2006), o que explica os valores elevados de AS observados em São Paulo e no leste dos EUA. No entanto, a quantidade relativa de carbono elementar (CE) ou de *black carbon* (BC) na massa do $MP_{2.5}$ em São Paulo é pelo menos quatro vezes a do leste dos EUA. O estudo realizado por Martins Pereira et al. (2017) em São Paulo, mostrou que o CE contribuiu com cerca de 16% da massa do $MP_{2.5}$, enquanto essa mesma contribuição é de cerca de 4% no leste dos EUA, como relatado por Bell et al. (2007). Os estudos conduzidos em São Paulo por Castanho and Artaxo (2001) e Andrade et al. (2012b) relataram que o BC contribuiu com cerca de 25% e 36% da massa do $MP_{2.5}$, respectivamente, enquanto essa mesma contribuição é de cerca de 3% na costa leste dos EUA, como descrito no estudo conduzido por Castanho et al. (2005). Como o CE e o BC são, em geral, absorventes (não refletem radiação solar), o AS do aerossol em São Paulo é menor do que o AS do aerossol no leste dos EUA.

Figura 4.2: Média de AS para cada bin de FMF e TAP nas estações GSFC (esquerda) e São Paulo (direita) da rede AERONET. Apenas os bins com pelo menos cinco observações foram considerados nos gráficos.

Além da correlação entre os dados de POA obtidos da rede AERONET e os obtidos via satélite, outra forma de se avaliar o desempenho de um algoritmo de aquisição de dados de POA via satélite é por meio do envelope de erro esperado (EEE). No caso do algoritmo MAIAC, o EEE é definido como $\pm (0,05 + 0,05POA)$ e espera-se que pelo menos 66% dos dados de POA obtidos por meio do algoritmo MAIAC, o que corresponde a 2/3 dos dados ou aproximadamente 1 desvio-padrão, caiam dentro do EEE (Martins et al., 2017). Na área urbana da América do Sul, que contempla as cidades de São Paulo e Buenos Aires, menos de 66% dos dados de POA do MODIS-MAIAC caíram dentro do EEE: 57,7% e 54,7% para os satélites Aqua e Terra, respectivamente (Martins et al., 2017). Realizando os cálculos considerando apenas os dados da Cidade de São Paulo, embora tenham sido encontradas correlações relativamente altas entre a POA do MODIS-MAIAC e a POA da AERONET ($R_{AQUA}=0.90$ e $R_{TERRA}=0.82$), as proporções de dados de POA do MODIS-MAIAC que caíram dentro do EEE foram iguais a 45% e 38%, para os satélites Aqua e Terra, respectivamente, como ilustrado na Figura 4.3.

Esses resultados evidenciam que apesar de os dados de POA do MODIS-MAIAC apresentarem um bom desempenho em áreas com ampla cobertura vegetal, o mesmo não é observado para as áreas urbanas na América do Sul.

Figura 4.3: Comparação entre os dados de POA da estação São Paulo da rede AERONET com os dados de POA obtidos do algoritmo MAIAC para os satélites Aqua (esquerda) e Terra (direita). A linhas pontilhadas vermelhas e cinza representam o EEE ($EEE = \pm (0.05 + 0.05 \times POA)$) e a linha $y=x$, respectivamente. A linha contínua em preto representa a linha da regressão entre os dados de POA da AERONET e do MAIAC. No canto superior esquerdo temos que: R é o coeficiente da correlação linear de Pearson entre os dados de POA da AERONET e do MAIAC, n é o número de pares (POA-MAIAC; POA-AERONET) e “dentro do EEE” é a proporção de dados de POA do MODIS-MAIAC que caíram dentro do EEE.

4.3 Modelo proposto para a predição das concentrações de $MP_{2.5}$ na RMSP

Como demonstrado na seção 4.2, a POA obtida da versão 6 dos dados do MODIS-MAIAC não é uma preditora das concentrações de $MP_{2.5}$ e MP_{10} medidas na superfície da RMSP. Apesar disso, o MLM ainda pode ser utilizado para estimar a concentração de $MP_{2.5}$ para cada célula da grade esquematizada na Figura 3.2 no período do estudo. Dessa forma, foi proposta uma metodologia baseada apenas nos estágios 1 e 2 da *Metodologia de Harvard*, com algumas alterações no primeiro estágio.

4.3.1 Primeiro estágio

Partindo dos modelos esquematizados na Tabela 3.7, tanto para os dados diurnos quanto para os dados diários, realizamos os procedimentos descritos na seção 3.3.3.1 para obter os respectivos modelos finais apresentados nas Tabelas 4.5 e 4.6.

Referente às hipóteses apresentadas na seção 3.3.3 podemos dizer que:

1. quando comparamos os modelos finais ajustados a partir do conjunto de variáveis explicativas E2 com os modelos ajustados a partir do conjunto de variáveis explicativas E1 ($E1A1 \times E2A1$ e $E1A2 \times E2A2$) observamos que, além das variáveis meteorológicas (nos modelos E2A1 e E2A2) e da densidade viária (no modelo E2A1), as variáveis indicadoras da ocorrência de feriado e dos dias da semana, mas não a interação entre elas, melhorou o poder de predição do $MP_{2.5}$, para as concentrações diárias. Os valores de AIC dos modelos finais ajustados a partir do conjunto de variáveis E2 são cerca de 0,01% menores do que os valores de AIC dos modelos finais ajustados a partir do conjunto de variáveis E1. Para as concentrações diurnas, além das variáveis meteorológicas (nos modelos E2A1 e E2A2) e da densidade viária (no modelo E2A1), as variáveis indicadoras da ocorrência de feriado e dos dias da semana, bem como a interação entre elas, melhorou o poder de predição do $MP_{2.5}$ nos modelos E2A2, enquanto as variáveis indicadora da ocorrência de feriado e dos dias da semana, mas não a interação entre elas, melhorou o poder de predição do $MP_{2.5}$ nos modelos E2A1. Os valores de AIC dos modelos finais ajustados a partir do conjunto de variáveis E2 são cerca de 0,04% menores do que os valores de AIC dos modelos finais ajustados a partir do conjunto de variáveis E1.
2. quando comparamos os modelos finais ajustados a partir do conjunto de variáveis explicativas E3 com os modelos ajustados a partir do conjunto de variáveis explicativas E1 ($E1A1 \times E3A1$ e $E1A2 \times E3A2$) observamos que, além das variáveis indicadoras da ocorrência de feriado e dos dias da semana, mas não a interação entre elas, das variáveis meteorológicas (nos modelos E3A1 e E3A2) e da densidade viária (no modelo E3A1),

a inclusão dos termos específicos (senos e cossenos) para capturar a sazonalidade das concentrações ao longo do tempo melhorou o poder de predição do $MP_{2.5}$, para as concentrações diárias. Para as concentrações diurnas, além das variáveis indicadoras da ocorrência de feriado e dos dias da semana, bem como a interação entre elas, das variáveis meteorológicas (nos modelos E3A1 e E3A2) e da densidade viária (no modelo E3A1), a inclusão dos termos específicos (senos e cossenos) para capturar a sazonalidade das concentrações ao longo do tempo melhorou o poder de predição do $MP_{2.5}$. Os valores de AIC dos modelos finais ajustados a partir do conjunto de variáveis E3 são cerca de 0,2% menores do que os valores de AIC dos modelos finais ajustados a partir do conjunto de variáveis E1, tanto para os modelos ajustados com as concentrações diurnas quanto para os modelos ajustados com as concentrações diárias.

3. quando comparamos os modelos do tipo A1 com os do tipo A2 (E1A1 x E1A2, E2A1 x E2A2 e E3A1 x E3A2), observamos que os modelos do tipo A2 apresentaram valores de AIC cerca de 5% menores do que os valores de AIC dos modelos do tipo A1, tanto para os dados diurnos quanto para os dados diários, indicando que os modelos que consideram a dependência espacial e temporal, segundo o critério de informação AIC, têm melhor poder de predição do que aqueles que levam em consideração apenas a dependência espacial.

Tabela 4.5: Estimativas (erros-padrão) dos parâmetros dos modelos finais para as médias diurnas das concentrações de $MP_{2.5}$ obtidos a partir dos modelos esquematizados na Tabela 3.7.

	E1A1	E2A1	E3A1	E1A2	E2A2	E3A2
(Intercepto)	17,366 (0,741)	14,932 (0,791)	7,278 (0,838)	15,009 (0,830)	12,394 (0,879)	4,912 (0,652)
Temperatura	0,055 (0,023)	0,060 (0,022)	0,318 (0,025)	0,141 (0,022)	0,147 (0,022)	0,407 (0,019)
Umidade	-0,049 (0,004)	-0,049 (0,004)	-0,023 (0,004)	-0,024 (0,004)	-0,025 (0,004)	
Precipitação		-0,007 (0,004)		-0,012 (0,003)	-0,012 (0,003)	-0,010 (0,003)
Velocidade do Vento	-0,129 (0,030)	-0,136 (0,029)	-0,131 (0,029)	-0,292 (0,032)	-0,298 (0,032)	-0,305 (0,032)
Densidade Viária	3,020 (0,109)	3,017 (0,109)	2,949 (0,109)			
Feriado		-3,200 (0,585)	2,272 (1,440)		1,287 (1,659)	2,367 (1,434)
Seg		2,547 (0,447)	2,788 (0,399)		2,820 (0,460)	2,820 (0,397)
Ter		3,301 (0,447)	3,589 (0,398)		3,624 (0,459)	3,629 (0,397)
Qua		3,446 (0,447)	3,624 (0,400)		3,679 (0,461)	3,659 (0,398)
Qui		3,463 (0,447)	3,657 (0,399)		3,672 (0,460)	3,711 (0,398)
Sex		3,456 (0,447)	3,646 (0,399)		3,675 (0,460)	3,668 (0,398)
Sáb		1,632 (0,447)	1,798 (0,396)		1,771 (0,457)	1,812 (0,395)
Feriado:Seg			-2,014 (0,151)		-6,271 (2,231)	-2,117 (0,148)
Feriado:Ter			0,857 (0,149)		-7,280 (2,304)	0,869 (0,148)
Feriado:Qua			-3,278 (0,168)		-5,045 (2,175)	-3,525 (0,161)
Feriado:Qui			-0,635 (0,148)		-4,310 (2,202)	-0,612 (0,147)
Feriado:Sex			-6,053 (1,937)		-5,062 (2,201)	-6,358 (1,929)
Feriado:Sáb			-6,605 (2,001)		-2,372 (2,717)	-6,722 (1,992)
$sen(2\pi \times t/365)$			-4,804 (1,887)			-4,953 (1,879)
$sen(2\pi \times 2t/365)$			-5,084 (1,912)			-5,255 (1,904)
$cos(2\pi \times t/365)$			-4,973 (1,910)			-5,062 (1,903)
$cos(2\pi \times 2t/365)$			-3,541 (2,355)			-3,398 (2,347)
σ^2	20,06	20,07	20,11	15,009	15,012	15,039
σ_u^2	32,5	30,36	22,94	33,315	31,06	22,99
σ_s^2				4,898	4,895	4,907
AIC	299094	298980	298493	285467	285353	284801

Tabela 4.6: Estimativas (erros-padrão) dos parâmetros dos modelos finais para as médias diárias das concentrações de $MP_{2.5}$ obtidos a partir dos modelos esquematizados na Tabela 3.7.

	E1A1	E2A1	E3A1	E1A2	E2A2	E3A2
(Intercepto)	23,956 (0,756)	22,415 (0,823)	15,338 (0,857)	15,529 (0,595)	13,956 (0,686)	7,124 (0,902)
Temperatura	-0,112 (0,024)	-0,108 (0,024)	0,148 (0,026)	0,088 (0,021)	0,092 (0,021)	0,359 (0,025)
Umidade	-0,062 (0,004)	-0,062 (0,004)	-0,039 (0,004)			0,018 (0,005)
Precipitação				-0,006 (0,002)	-0,006 (0,002)	-0,005 (0,002)
Velocidade do Vento	-0,431 (0,031)	-0,433 (0,031)	-0,421 (0,031)	-0,425 (0,033)	-0,427 (0,033)	-0,443 (0,033)
Densidade Viária	1,341 (0,100)	1,336 (0,100)	1,218 (0,100)			
Feriado		-2,533 (0,656)	-1,560 (0,579)		-2,722 (0,683)	-1,569 (0,584)
Seg		1,240 (0,501)	1,236 (0,441)		1,249 (0,522)	1,242 (0,445)
Ter		2,051 (0,501)	2,076 (0,441)		2,090 (0,522)	2,114 (0,445)
Qua		2,170 (0,501)	2,156 (0,441)		2,205 (0,522)	2,185 (0,445)
Qui		2,305 (0,501)	2,295 (0,441)		2,339 (0,522)	2,327 (0,445)
Sex		2,267 (0,501)	2,267 (0,441)		2,283 (0,522)	2,283 (0,445)
Sáb		1,130 (0,501)	1,148 (0,441)		1,126 (0,522)	1,148 (0,445)
Feriado:Seg						
Feriado:Ter						
Feriado:Qua						
Feriado:Qui						
Feriado:Sex						
Feriado:Sáb						
$sen(2\pi \times t/365)$			-1,962 (0,170)			-2,257 (0,171)
$sen(2\pi \times 2t/365)$			1,111 (0,167)			1,164 (0,169)
$cos(2\pi \times t/365)$			-3,616 (0,185)			-4,247 (0,185)
$cos(2\pi \times 2t/365)$						
σ^2	16,38	16,38	16,41	11,83	11,83	11,85
σ_u^2	39,48	38,56	29,72	43,11	42,14	30,48
σ_s^2				4,32	4,32	4,35
AIC	287817	287783	287335	272763	272729	272142

No que diz respeito à etapa de validação dos modelos finais, as Tabelas 4.7 e 4.8 mostram as medidas de desempenho, R^2 e $REQMP$, para os modelos apresentados nas Tabelas 4.5 e 4.6.

O desempenho dos modelos que consideram as duas fontes de correlação (modelos A2) foi ligeiramente superior ao desempenho dos modelos que, assim como na *Metodologia de Harvard*, consideram apenas a correlação espacial (modelos A1), tanto para os dados diurnos quanto para os dados diários. Globalmente, os modelos do tipo A2, para as concentrações diurnas (diárias), apresentaram valores de R^2 da ordem de 0,62 (0,71) e $REQMP$ da ordem de 4,6 (4,1) $\mu g/m^3$. Dentro de cada tipo de modelo, comparando os conjuntos de variáveis explicativas, os desempenhos dos modelos E1 e E2 podem ser considerados equivalentes e ligeiramente superiores ao desempenho dos modelos E3, tanto para os dados diurnos quanto para os diários.

Analisando as medidas de desempenho dos modelos para cada estação de monitoramento, a estação Pico do Jaraguá foi a única que apresentou valores de R^2 inferiores a 0,50, tanto para os dados diurnos quanto para os dados diários. A estação Pico do Jaraguá (Cerqueira César) foi a que apresentou o menor (maior) valor de R^2 , para as concentrações diurnas e diárias, em todos os modelos. A estação Marginal Tietê - Ponte dos Remédios (Cerqueira César) foi a que apresentou maior (menor) valor de $REQMP$, para as concentrações diurnas. De forma similar, a estação Marginal Tietê - Ponte dos Remédios (Carapicuíba) foi a que apresentou maior (menor) valor de $REQMP$, para as concentrações diárias.

Tabela 4.7: Medidas de desempenho, para cada estação de monitoramento e no geral, dos modelos finais apresentados na Tabela 4.5

Estação	E1A1		E2A1		E3A1		E1A2		E2A2		E3A2	
	R^2	REQMP	R^2	REQMP	R^2	REQMP	R^2	REQMP	R^2	REQMP	R^2	REQMP
Cerqueira César	0,8869	2,4459	0,8871	2,4431	0,8828	2,4531	0,8920	2,3281	0,8923	2,3266	0,8881	2,3410
Diadema	0,8286	3,0262	0,8286	3,0434	0,8265	3,0315	0,8358	2,9615	0,8364	2,9621	0,8340	2,9668
São Caetano do Sul	0,8244	3,4680	0,8247	3,4651	0,8221	3,4438	0,8246	3,4379	0,8247	3,4370	0,8205	3,4249
Santo André - Paço Municipal	0,8194	4,0527	0,8200	4,0537	0,8200	4,1139	0,8249	3,1782	0,8252	3,1788	0,8242	3,2557
Interlagos	0,8147	4,2014	0,8136	4,2030	0,8089	4,2113	0,8211	3,8734	0,8203	3,8742	0,8164	3,8710
Capão Redondo	0,8124	3,4532	0,8117	3,4581	0,8140	3,4312	0,8155	3,5138	0,8148	3,5181	0,8168	3,4869
Carapicuíba	0,8084	2,9535	0,8088	2,9501	0,8087	2,9420	0,8029	3,0723	0,8031	3,0715	0,8031	3,0541
Santo Amaro	0,8083	2,8376	0,8079	2,8385	0,8072	2,8327	0,8092	2,8982	0,8091	2,8950	0,8088	2,8767
Moóca	0,7902	3,5101	0,7890	3,5186	0,7871	3,5517	0,7909	3,4408	0,7900	3,4479	0,7877	3,4857
São Bernardo do Campo-Paulicéia	0,7702	3,6066	0,7700	3,6082	0,7660	3,6340	0,7739	3,6330	0,7738	3,6328	0,7703	3,6507
Nossa Senhora do Ó	0,7603	3,9992	0,7608	3,9955	0,7630	4,0255	0,7619	4,2673	0,7623	4,2641	0,7642	4,2851
Santana	0,7524	3,4565	0,7521	3,4587	0,7557	3,4343	0,7489	3,4991	0,7489	3,4990	0,7522	3,4708
Parque D. Pedro II	0,7458	4,5860	0,7464	4,5806	0,7515	4,4906	0,7413	3,4473	0,7420	3,4438	0,7478	3,3922
Santo André - Capuava	0,7421	3,1539	0,7423	3,1497	0,7402	3,1283	0,7437	3,3575	0,7438	3,3492	0,7411	3,3084
Congonhas	0,7360	4,2476	0,7359	4,2503	0,7329	4,3301	0,7383	4,2444	0,7383	4,2470	0,7360	4,3167
Marg. Tietê - Ponte dos Remédios	0,7018	9,6587	0,7025	9,6525	0,7031	9,5911	0,7029	9,5504	0,7036	9,5348	0,7047	9,4907
São Bernardo do Campo-Centro	0,7011	4,0742	0,7003	4,0794	0,6939	4,1310	0,7021	4,1874	0,7010	4,1930	0,6954	4,2472
Guarulhos - Pimentas	0,6949	4,0457	0,6952	4,0434	0,6934	4,0541	0,6975	4,0492	0,6975	4,0481	0,6965	4,0588
Taboão da Serra	0,6910	4,5712	0,6907	4,5811	0,6974	4,6883	0,7058	4,6785	0,7054	4,6906	0,7182	4,7426
Pinheiros	0,6900	5,5445	0,6891	5,5470	0,6814	5,5305	0,6957	5,5789	0,6948	5,5792	0,6870	5,5505
Osasco	0,6841	6,3427	0,6853	6,3382	0,6876	6,3636	0,6802	6,3880	0,6812	6,3794	0,6828	6,4049
Guarulhos - Paço Municipal	0,6770	4,1442	0,6771	4,1452	0,6786	4,1742	0,6673	4,2790	0,6676	4,2824	0,6696	4,3157
Itaim Paulista	0,6766	3,9441	0,6761	3,9472	0,6767	3,9374	0,6744	4,1622	0,6739	4,1640	0,6744	4,1413
Ibirapuera	0,6428	4,5592	0,6412	4,5689	0,6434	4,5646	0,6361	4,6605	0,6350	4,6673	0,6372	4,6660
Parelheiros	0,6355	8,5493	0,6356	8,5574	0,6417	8,5031	0,6438	7,1120	0,6439	7,1223	0,6462	7,1035
Cid. Universitária - USP	0,6056	5,2036	0,6043	5,2119	0,5937	5,2935	0,5950	5,2845	0,5941	5,2906	0,5856	5,3585
Mauá	0,5171	5,1497	0,5160	5,1587	0,5028	5,2428	0,5104	5,3529	0,5096	5,3617	0,4957	5,4553
Pico do Jaraguá	0,4439	5,8273	0,4446	5,8230	0,4521	5,7935	0,4434	5,7198	0,4440	5,7163	0,4505	5,6680
Todas	0,6022	4,7482	0,6020	4,7499	0,6013	4,7536	0,6230	4,6207	0,6230	4,6209	0,6222	4,6257

Tabela 4.8: Medidas de desempenho, para cada estação de monitoramento e no geral, dos modelos finais apresentados na Tabela 4.6

Estação	E1A1		E2A1		E3A1		E1A2		E2A2		E3A2	
	R^2	REQMP	R^2	REQMP	R^2	REQMP	R^2	REQMP	R^2	REQMP	R^2	REQMP
Cerqueira César	0,9087	3,2960	0,9089	3,2961	0,9060	3,3059	0,9146	3,0855	0,9147	3,0859	0,9116	3,1382
Santo André - Paço Municipal	0,8866	2,9781	0,8868	2,9770	0,8884	2,9901	0,8943	2,6803	0,8944	2,6808	0,8934	2,7420
Diadema	0,8760	2,9662	0,8763	2,9630	0,8760	2,8849	0,8825	2,9331	0,8827	2,9306	0,8827	2,8432
Moóca	0,8761	2,7533	0,8759	2,7571	0,8751	2,8263	0,8756	2,7384	0,8754	2,7416	0,8743	2,8282
Carapicuíba	0,8757	2,7527	0,8757	2,7522	0,8755	2,7321	0,8731	2,7689	0,8731	2,7681	0,8725	2,7466
São Caetano do Sul	0,8736	3,2023	0,8737	3,2015	0,8739	3,1631	0,8779	3,0779	0,8779	3,0773	0,8760	3,0587
Interlagos	0,8694	3,8632	0,8692	3,8629	0,8669	3,8640	0,8719	3,7061	0,8717	3,7054	0,8701	3,6748
Santo Amaro	0,8595	2,8400	0,8593	2,8396	0,8593	2,8456	0,8612	2,9957	0,8611	2,9948	0,8613	2,9681
Capão Redondo	0,8543	3,6451	0,8540	3,6465	0,8541	3,6178	0,8575	3,6670	0,8573	3,6673	0,8567	3,6289
São Bernardo do Campo-Paulicéia	0,8534	3,2037	0,8534	3,2036	0,8497	3,2302	0,8553	3,2893	0,8553	3,2892	0,8520	3,3040
Osasco	0,8455	5,2783	0,8460	5,2755	0,8459	5,3327	0,8441	5,2712	0,8444	5,2695	0,8441	5,3100
Nossa Senhora do Ó	0,8376	3,5822	0,8377	3,5830	0,8375	3,6498	0,8388	3,7068	0,8388	3,7073	0,8384	3,7631
Santana	0,8253	3,0519	0,8252	3,0528	0,8268	3,0500	0,8228	3,0715	0,8228	3,0721	0,8242	3,0717
Parque D. Pedro II	0,8158	3,9562	0,8159	3,9551	0,8174	3,8995	0,8119	3,2918	0,8119	3,2920	0,8137	3,2930
Marg. Tietê - Ponte dos Remédios	0,8133	8,8271	0,8136	8,8234	0,8146	8,7788	0,8167	8,5752	0,8169	8,5713	0,8171	8,4377
Guarulhos - Pimentas	0,8009	4,7948	0,8010	4,7948	0,8013	4,8235	0,8014	4,8299	0,8014	4,8305	0,8017	4,8689
Santo André - Capuava	0,7974	3,0035	0,7974	3,0025	0,7965	2,9804	0,7981	3,0929	0,7982	3,0912	0,7966	3,0427
Pinheiros	0,7968	4,9102	0,7966	4,9111	0,7900	4,9223	0,8033	4,8715	0,8031	4,8719	0,7944	4,8743
Congonhas	0,7880	4,4469	0,7881	4,4469	0,7867	4,4922	0,7887	4,4046	0,7888	4,4045	0,7877	4,4389
Taboão da Serra	0,7826	3,5272	0,7824	3,5308	0,7864	3,5902	0,8055	3,3963	0,8054	3,3988	0,8017	3,5326
São Bernardo do Campo-Centro	0,7828	3,4431	0,7824	3,4465	0,7796	3,4920	0,7874	3,4632	0,7871	3,4657	0,7830	3,5222
Cid. Universitária - USP	0,7809	3,9105	0,7806	3,9135	0,7821	3,9367	0,7777	3,9504	0,7774	3,9534	0,7576	4,1189
Itaim Paulista	0,7688	3,9807	0,7686	3,9823	0,7692	3,9766	0,7664	3,9451	0,7663	3,9465	0,7664	3,9485
Mauá	0,7634	3,5061	0,7631	3,5090	0,7567	3,5721	0,7624	3,5209	0,7622	3,5230	0,7536	3,6024
Guarulhos - Paço Municipal	0,7447	3,8978	0,7448	3,8983	0,7449	3,9048	0,7295	4,0162	0,7296	4,0164	0,7283	4,0253
Parelheiros	0,7370	7,0063	0,7368	7,0115	0,7390	7,0934	0,7495	5,9460	0,7494	5,9495	0,7485	6,1375
Ibirapuera	0,7042	4,1363	0,7037	4,1399	0,7041	4,1383	0,6989	4,1809	0,6986	4,1832	0,6986	4,1843
Pico do Jaraguá	0,4942	6,5156	0,4947	6,5123	0,5046	6,4640	0,4968	6,2683	0,4972	6,2661	0,5069	6,2185
Todas	0,6978	4,2786	0,6977	4,2788	0,6965	4,2876	0,7156	4,1502	0,7156	4,1502	0,7136	4,1650

Esses resultados mostram que a diferença de desempenho dos modelos ajustados é relativamente pequena, e por isso optamos por seguir apresentando os resultados dos modelos E1A1 e E1A2, pois esses foram os que apresentaram o melhor desempenho independente da estrutura de correlação considerada ($A1 \times A2$), tanto para os dados diurnos quanto para os dados diários.

As Figuras 4.4 - 4.7 apresentam a comparação entre as médias diárias observadas das concentrações de $MP_{2.5}$ e as médias diárias preditas pelos modelos E1A1 e E1A2, para os dados das estações Cerqueira César e Pico do Jaraguá, que foram as estações que apresentaram o melhor e o pior desempenho, respectivamente. Os gráficos para as demais estações de monitoramento, quando considerados os dados diários, bem como os gráficos referentes aos dados diurnos, para ambos os modelos, estão nos Apêndices G, H I e J, respectivamente.

Nas Figuras 4.4 - 4.7, as linhas amarela e vermelha representam as predições condicionais e marginais, que são os valores de $MP_{2.5}$ preditos pelos modelos quando os efeitos aleatórios u_t são e não são considerados nas predições, respectivamente. Note que independente do modelo, as predições condicionais são as que melhor acompanham a série de dados de $MP_{2.5}$, representada em preto, indicando que além de considerar as variáveis meteorológicas e de uso e ocupação da solo nos modelos de predição, os efeitos aleatórios desempenham papel essencial nas predições de $MP_{2.5}$ na RMSP.

Figura 4.4: Médias diárias da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A1 para a estação de monitoramento Cerqueira César em 2012-2017.

Figura 4.5: Médias diárias da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A2 para a estação de monitoramento Cerqueira César em 2012-2017.

Figura 4.6: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Pico do Jaraguá em 2012-2017.

Figura 4.7: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Pico do Jaraguá em 2012-2017.

4.3.2 Segundo estágio

No segundo estágio, os modelos finais E1A1 e E1A2 apresentados nas Tabelas 4.5 e 4.6, foram utilizados para construir os mapas com as médias anual e sazonais das concentrações diurnas (Figuras 4.8 e 4.9) e diárias (Figuras 4.10 e 4.11) previstas de $MP_{2.5}$. Além disso, são apresentados os respectivos mapas com a média anual dos desvios-padrão dos erros de predição. Para uma melhor visualização os mapas foram construídos com escalas diferentes.

Os mapas construídos na mesma escala encontram-se no Apêndice K.

Como a densidade viária foi uma das variáveis significativas nos modelos de predição do tipo A1, tanto para as concentrações diurnas quanto para as concentrações diárias, os mapas apresentados nas Figuras 4.8 e 4.10 discriminam muito bem as altas concentrações associadas às vias com maior densidade viária concentradas na região delimitada pela Zona de Rodízio e às vias que dão acesso a essa região. O mesmo não foi observado nos modelos de predição do tipo A2, tanto para as concentrações diurnas quanto para as concentrações diárias, nos quais apenas as variáveis meteorológicas foram significativas nos modelos de predição finais (Figuras 4.9 e 4.11).

Referente à sazonalidade das concentrações, em ambos os modelos, as concentrações preditas são maiores no inverno e na primavera e menores no outono e verão (Apêndice K).

Quanto aos desvios-padrão dos erros de predição, no geral, observa-se que na região central, onde estão localizadas a maioria das estações de monitoramento, os desvios-padrão dos erros de predição são menores, em todos os modelos. Além disso, os desvios-padrão dos erros de predição nos modelos do tipo A2 são menores do que os dos modelos do tipo A1, indicando que as estimativas obtidas dos modelos que consideram as correlações espacial e temporal são mais precisas.

Como observado na Figura 3.1, a densidade das estações que fazem o monitoramento do $MP_{2.5}$ e/ou do MP_{10} é maior na Cidade de São Paulo e no seu entorno. O mesmo ocorre com a densidade das estações meteorológicas, exceto para as estações do DAEE (Figuras 3.6 e 3.7). Pensando na atual representatividade espacial dessas estações, consideramos que tanto os resultados das interpolações espaciais realizadas para as variáveis meteorológicas quanto as concentrações de $MP_{2.5}$ são mais representativas da Cidade de São Paulo do que da RMSP. Dessa forma, as estimativas das concentrações de $MP_{2.5}$ obtidas a partir dos modelos de predição apresentados na tese, do ponto de vista espacial, são mais representativas da Cidade de São Paulo do que da RMSP.

No Apêndice L apresentamos os mapas com as médias anual e sazonais das concentrações preditas de $MP_{2.5}$ apenas para a Cidade de São Paulo. Os respectivos mapas construídos na mesma escala encontram-se no Apêndice M.

Os mapas em formato *raster* (.tif) com a média anual das estimativas das concentrações diurnas e diárias de $MP_{2.5}$ para a RMSP e para a Cidade de São Paulo foram disponibilizadas em <https://repositorio.uspdigital.usp.br/>.

Figura 4.8: Médias anual (A) e sazonais (C-F) das concentrações diurnas de $MP_{2.5}$ preditas pelo modelo E1A1 para a RMSP em 2012-2017. (B) Média dos desvios-padrão dos erros de predição. Para uma melhor visualização os gráficos estão em escalas diferentes.

Figura 4.9: Médias anual (A) e sazonais (C-F) das concentrações diurnas de $MP_{2.5}$ preditas pelo modelo E1A2 para a RMSP em 2012-2017. (B) Média dos desvios-padrão dos erros de predição. Para uma melhor visualização os gráficos estão em escalas diferentes.

Figura 4.10: Médias anual (A) e sazonais (C-F) das concentrações diárias de $MP_{2.5}$ preditas pelo modelo E1A1 para a RMSP em 2012-2017. (B) Média dos desvios-padrão dos erros de predição. Para uma melhor visualização os gráficos estão em escalas diferentes.

Figura 4.11: Médias anual (A) e sazonais (C-F) das concentrações diárias de $MP_{2.5}$ preditas pelo modelo E1A2 para a RMSP em 2012-2017. (B) Média dos desvios-padrão dos erros de predição. Para uma melhor visualização os gráficos estão em escalas diferentes.

5. Discussão

Os habitantes de áreas urbanas, como os de São Paulo, são diariamente expostos à poluentes no ar em níveis superiores aos preconizados pela OMS. A associação negativa entre as concentrações de $MP_{2.5}$ e os desfechos relacionadas à saúde humana, como problemas respiratórios e cardiovasculares, já é bem conhecida (World Health Organization, 2006). As concentrações dos poluentes podem ser obtidas por meio de dados de estações em redes de monitoramento da qualidade do ar, que em geral, são poucas e mal distribuídas no espaço, e são representativas apenas dos habitantes que vivem ou trabalham no entorno dessas redes. Assim, se faz necessário o desenvolvimento de metodologias que permitam estimar as concentrações de $MP_{2.5}$ nas regiões onde faltam redes de monitoramento da qualidade do ar.

Nesta tese, construímos um modelo espaço-temporal para a predição da concentração de $MP_{2.5}$ na RMSP, no período compreendido entre 2012 e 2017, baseado na metodologia proposta pelos pesquisadores da Faculdade de Saúde Pública de Harvard. Essa metodologia, que combina um modelo estatístico (MLM) com dados de estações de monitoramento, dados de POA obtidos de satélites, variáveis meteorológicas e variáveis de uso e ocupação do solo, tem sido utilizada de forma satisfatória para realizar a predição das concentrações $MP_{2.5}$ e MP_{10} em diferentes regiões do mundo (Lee et al., 2011; Kloog et al., 2011, 2012; Chudnovsky et al., 2014b; Stafoggia et al., 2017; de Hoogh et al., 2018; Xiao et al., 2017; Just et al., 2015; Kloog et al., 2015).

A principal contribuição dessa metodologia foi a utilização dos dados de POA obtidos de satélite em conjunto com o MLM. Apesar de a POA ser uma medida associada à coluna atmosférica e a concentração de $MP_{2.5}$ ser uma medida realizada próxima à superfície, aquela tem sido utilizada como uma aproximação para esta em diferentes regiões do mundo. Em um estudo de revisão, Chu et al. (2016) incluíram 116 estudos que utilizaram algum modelo para a predição do $MP_{2.5}$ na superfície a partir de medidas de POA obtidas por satélite, e constataram que quatro tipos de modelos foram os mais utilizados para a predição das concentrações do $MP_{2.5}$: Regressão Linear Múltipla (RLM), Modelo Linear Misto (MLM), Modelo de Transporte Químico (MTQ) e Regressão Geograficamente Ponderada (RGP). O MLM (RLM) foi o que apresentou melhor (pior) desempenho. O MTQ (RGP) apresentou melhor predição em escala global (regional). No entanto, o MLM, que é o mesmo modelo utilizado na *Metodologia de Harvard*, apresenta uma vantagem adicional que é uma abordagem para o tratamento dos dados faltantes de POA, que em geral ocorrem sobre superfícies muito refletoras (desertos, superfícies cobertas por neve ou com pouca cobertura vegetal) ou quando há presença de nuvens.

Como apresentado na seção 4.1, encontramos resultados inesperados ao aplicar a *Metodologia de Harvard* para a RMSP, utilizando dados de concentração de $MP_{2.5}$, dados de POA

obtidos do MODIS-MAIAC, dados meteorológicos e a densidade viária. O sinal associado à POA obtida do MODIS-MAIAC foi negativo para a metade dos modelos ajustados, diferente do observado nos demais estudos que haviam aplicado essa mesma metodologia. A partir daí, iniciamos diversas investigações na tentativa de entender o resultado encontrado.

A primeira delas foi aplicar a mesma estratégia utilizada por [Stafoggia et al. \(2017\)](#) e [de Hoogh et al. \(2018\)](#), que aplicaram a *Metodologia de Harvard* em dados da Itália e da Suíça, respectivamente. Em ambos os trabalhos, os autores removeram das análises os pares (POA, $MP_{2.5}$) considerados conflitantes: valores altos/baixos de POA com valores baixos/altos de $MP_{2.5}$. Nenhuma justificativa ou algum tipo de discussão sobre o motivo de esses pares serem conflitantes no contexto dos respectivos estudos foi apresentada pelos autores.

Nos dados da tese, em princípio, achamos que os pares conflitantes poderiam estar associados ao período de queimadas que ocorrem na região Norte e Central do Brasil ou mesmo no interior do estado de São Paulo. Nesse período poderia existir um desacordo entre os dados medidos na superfície e os dados mensurados pelo satélite. As estações de monitoramento na superfície não conseguem mensurar as concentrações em camadas mais elevadas da atmosfera trazidas pelo transporte de outras regiões, enquanto que o satélite consegue. No entanto, os dados não confirmaram essa hipótese. E mesmo após a exclusão dos pares conflitantes não foi observada nenhuma melhora sistemática na relação entre a POA e o $MP_{2.5}$ no período estudado.

Pensamos em utilizar os dados do instrumento CALIOP a bordo do satélite CALIPSO. Com os dados fornecidos por esse instrumento seria possível obter uma estimativa da quantidade de aerossóis presentes acima da CLP. Dessa forma, em vez de utilizarmos a POA obtida do MODIS-MAIAC como preditora das concentrações de $MP_{2.5}$, poderíamos utilizar a diferença obtida entre os dados do MODIS-MAIAC e os dados obtidos do CALIPSO e teríamos uma aproximação da quantidade de aerossol presente mais próximo à superfície para relacionar com os dados da CETESB: $MP_{CETESB} \sim POA_{MODIS-MAIAC} - POA_{CALIPSO}$. No entanto, as resoluções temporal e espacial dos dados do CALIPSO são bastante limitadas para o objetivo da tese. A resolução temporal do CALIPSO é de cerca de 16 dias. Quanto à resolução espacial, como a região de interesse deste estudo é relativamente pequena, apenas uma rota de passagem desse satélite cai dentro da área de interesse, o que limita a resolução espacial.

Também pensamos em utilizar os dados de POA obtidos da estação São Paulo da rede AERONET. Porém, a primeira limitação dessa alternativa é que, como se trata de uma medida associada a um único ponto dentro da região de interesse, estaríamos assumindo que a POA é homogênea em toda a RMSP, o que provavelmente não é verdade. A segunda limitação está na ausência dos dados de POA da estação São Paulo dentro do período do presente estudo, devido à calibração do fotômetro (Figura 5.1).

Figura 5.1: Valores de POA da estação São Paulo da rede AERONET em 2012-2017.

Apesar de já ter sido utilizada de forma satisfatória como preditora das concentrações de $MP_{2.5}$ em outros estudos, antes de utilizar os dados de POA obtidos de satélite em estudos epidemiológicos e de qualidade do ar em regiões específicas, sua aplicabilidade deve ser avaliada (Hersey et al., 2015). Na seção 4.2.1 apresentamos os resultados obtidos da correlação entre os dados do MODIS-MAIAC e as concentrações de $MP_{2.5}$ e MP_{10} na RMSP. Ao contrário do que é observado na maioria das regiões onde dados de POA obtidos de satélites já foram utilizados, encontramos correlações negativas e/ou fracas entre os dados do MODIS-MAIAC e as concentrações de $MP_{2.5}$ e MP_{10} medidos na superfície, mesmo levando em conta os efeitos da ACLP, da UR e das estações do ano, fatores importantes e que influenciam a relação POA-MP (seção 4.2.2).

Uma revisão de literatura mais detalhada mostrou que relações fracas e/ou negativas entre dados de POA obtidos de satélite e as concentrações de $MP_{2.5}$ e MP_{10} na superfície já haviam sido reportadas na RMSP por Natali (2008), como contextualizado na seção 3.3.2, e em outras regiões do mundo, principalmente na região oeste dos EUA.

Toth et al. (2014) estudaram a relação POA- $MP_{2.5}$ nos EUA utilizando dados de POA dos instrumentos MODIS (coleção 5.1 e algoritmo *Dark Target*), MISR e CALIOP entre 2008 e 2009. Eles descobriram que a correlação entre os dados de POA desses instrumentos com os dados de $MP_{2.5}$ medidos na superfície da região leste dos EUA é o dobro da observada na região oeste (0,2 vs. 0,4).

Engel-Cox et al. (2004), utilizando dados de POA da coleção 5 do MODIS, e Paciorek et al. (2008), utilizando o produto *Aerosol/Smoke* do *Geostationary Operational Environmental Satellite* (GOES-12), encontraram valores de correlação com padrões semelhantes: 0,6-0,8 (leste dos EUA) e 0,2-0,4 (oeste dos EUA).

Da mesma forma, Zhang et al. (2009) mostraram a diminuição nas correlações entre as concentrações horárias de $MP_{2.5}$ e os dados de POA do MODIS (coleção 5) do leste ($\sim 0,6$) para o oeste dos EUA ($\sim 0,2$).

Lee et al. (2016) encontraram um valor de correlação igual a 0,28 entre as concentrações diárias de $MP_{2.5}$ e os dados de POA do MODIS (coleção 6 e algoritmo *Deep Blue*) na Califórnia, EUA.

Loría-Salazar et al. (2016) atribuem as maiores correlações observadas no leste dos EUA à mistura bastante uniforme de sulfatos e aerossóis orgânicos encontrada na poluição dessa região, à refletância homogênea da superfície mais escura, à baixa elevação e à física menos complicada da camada limite nessa região.

Na Europa, Koelemeijer et al. (2006) também relataram baixas correlações entre as concentrações de $MP_{2.5}$ e MP_{10} (concentrações horárias e diárias) e a POA obtida do MODIS (coleção 4).

Na África do Sul, Hersey et al. (2015), com base em um conjunto de dados de dez anos (2000-2009), estudaram a relação POA-($MP_{2.5}$ e MP_{10}) nas cinco principais áreas metropolitanas e também encontraram correlações fracas e negativas entre as concentrações mensais de $MP_{2.5}$ e MP_{10} e os dados de POA dos sensores MODIS e MISR.

As justificativas apresentadas, principalmente pelos autores nos EUA, para as baixas correlações observadas entre as concentrações de $MP_{2.5}$ e os dados de POA obtidos de diferentes instrumentos e coleções foram: a variabilidade no tipo de aerossol (Engel-Cox et al., 2006; Lee et al., 2016), refletância da superfície (Engel-Cox et al., 2004, 2006; Paciorek et al., 2008; Zhang et al., 2009), o tipo de modelo de aerossol utilizado pelo MODIS na aquisição dos dados de POA (Engel-Cox et al., 2004, 2006) e a distribuição vertical dos aerossóis coluna atmosférica (Toth et al., 2014; Paciorek et al., 2008; Li et al., 2015).

Para a RMSP, os resultados apresentados na seção 4.2.3 mostraram que as baixas correlações encontradas entre as concentrações horárias, diurnas e diárias de $MP_{2.5}$ e MP_{10} medidos na superfície e os dados de POA do MODIS-MAIAC são devido principalmente ao desacordo entre o modelo de aerossol utilizado pelo algoritmo MAIAC na região onde encontra-se a RMSP e as propriedades físicas e químicas do aerossol dessa região.

A aquisição dos dados de POA via satélite depende fortemente do modelo assumido para representar as propriedades microfísicas do aerossol, bem como da refletância da superfície (Levy et al., 2007; de Almeida Castanho et al., 2008; Lyapustin et al., 2011b).

A versão 6 dos dados de POA do MODIS-MAIAC, atribui o mesmo modelo de aerossol para a costa leste dos EUA e para a América do Sul. No entanto, como visto na Figura 4.1, e já demonstrado no trabalho de Levy et al. (2007), o leste dos EUA é dominado por aerossóis não absorventes, enquanto a América do Sul (incluindo São Paulo) é dominada por aerossóis mais absorventes. Logo, para a RMSP, o MAIAC não poderia estar utilizando o mesmo modelo de aerossol da costa leste dos EUA.

Zhang et al. (2019) e Tao et al. (2019), que avaliaram as recuperações de POA obtidas por meio do algoritmo MAIAC sobre a China, também reportaram um desacordo entre o modelo de aerossol utilizado pelo MAIAC e as propriedades dos aerossóis locais. Na China, o MAIAC utiliza dois modelos de aerossol: um modelo “industrial” para a região leste, baseado nas características da estação Beijing da rede AERONET e o mesmo modelo atribuído para

a costa oeste dos EUA é utilizado sobre a região oeste. Zhang et al. (2019) concluem que o desempenho do MAIAC é melhor sobre as áreas urbanas e regiões com cobertura vegetal. No entanto, a má especificação do modelo de aerossol/a estimacão das propriedades da superfície sobre as regiões áridas e semi áridas, bem como sobre as regiões de água, induz grande incerteza na recuperação da POA nessas áreas. Já Tao et al. (2019) concluem que as propriedades do aerossol observadas em Beijing e na costa oeste dos EUA podem não ser aplicáveis às regiões leste e oeste da China, respectivamente.

Em versões anteriores dos algoritmos utilizados para a aquisição dos dados de POA a partir dos dados do MODIS, os modelos de aerossóis foram atualizados para melhor representar as propriedades físicas/ópticas de aerossóis locais. Em relação à América do Sul, na coleção 5 dos dados de POA do MODIS, foram especificados modelos sazonais de aerossol: um modelo de aerossol com $AS \sim 0,85$ foi atribuído para a área que compreende as regiões Centro-sul e Sudeste do Brasil, durante o inverno e a primavera, enquanto um modelo de aerossol de absorção moderada ($AS \sim 0,90$) foi atribuído para as demais regiões e durante o restante do ano (verão e outono) (ver Figura 3 de Levy et al. (2007)).

Os dados de POA do algoritmo MAIAC são relativamente novos e atualizações estão previstas. Na próxima versão dos dados do MAIAC, os dados de POA levarão em conta os efeitos da sazonalidade regional do aerossol, que ocorrem devido, por exemplo, aos efeitos causados por variações de umidade ou por períodos de queima de biomassa (Lyapustin et al., 2018).

Em relação à refletância da superfície, a RMSP é uma região desafiadora para a aquisição de dados de POA via satélite. Estudos de sensibilidade realizados por de Almeida Castanho et al. (2007) na Cidade do México e de Almeida Castanho et al. (2008) na Cidade de São Paulo, mostraram que a sensibilidade da aquisição da POA a partir dos dados do MODIS (versão 4) aumenta em função da refletância da superfície, sendo mais importante para valores menores de POA.

Nesse sentido, o algoritmo MAIAC melhorou a estimativa das relações espectrais da superfície, uma componente central para a aquisição dos dados de POA. Diferentemente do algoritmo *Dark Target*, que é baseado em taxas de refletância espectral pré-fixadas, o MAIAC utiliza uma abordagem de séries temporais e deriva dinamicamente essas taxas (Lyapustin et al., 2018), evitando assim suposições empíricas sobre propriedades da superfície (Martins et al., 2017). Então, em comparação ao produto MOD04 (resolução de 10 km), o MAIAC geralmente aumenta a precisão dos dados de POA sobre superfícies mais brilhantes, como mostrado nos estudos realizados nos EUA (Lyapustin et al., 2011a) e na China (Zhang et al., 2019). Porém, apesar de ter melhorado a estimacão das relações espectrais da superfície, a incerteza nos dados de POA obtidos do MAIAC aumenta à medida que a superfície se torna mais brilhante (Lyapustin et al., 2018).

O desempenho inferior dos dados de POA do MODIS-MAIAC sobre superfícies mais refletoras também pôde ser observado no estudo de validação dos dados de POA do MODIS-MAIAC na América do Sul. Martins et al. (2017) observaram uma menor correlacão entre

os dados de POA do MODIS-MAIAC e os da AERONET sobre superfícies brilhantes (áreas urbanas e áridas) em comparação à correlação observada sobre áreas com maior cobertura vegetal. As correlações na área urbana (São Paulo e Buenos Aires) foram as segundas mais baixas ($R = 0,69$ e $R = 0,64$, para os satélites Aqua e Terra, respectivamente). De forma similar, foram encontradas correlações mais baixas entre os dados de POA do MODIS-MAIAC e os da AERONET em locais da costa oeste do EUA, uma região de superfície mais brilhante com vegetação relativamente escassa ou inexistente (Lyapustin et al., 2011a).

Outro fator que desafia a aquisição de dados de POA na RMSP é a alta variabilidade de tipos de aerossóis. Além dos aerossóis locais, que são compostos principalmente por emissões veiculares e industriais, durante a época de queima de biomassa (agosto a setembro), a RMSP é afetada pelo transporte de longo alcance da fumaça vinda das regiões Norte e Centro-Oeste do Brasil, da Bolívia, do Paraguai, do Norte da Argentina e do interior do Estado de São Paulo (Mariano et al., 2010; de Miranda et al., 2017; Yamasoe et al., 2017; Vara-Vela et al., 2018).

Os trabalhos de Martins et al. (2009) e Yamasoe et al. (2017) mostraram que na RMSP o aerossol local é mais absorvente do que o transportado das áreas de queima de biomassa. Esses resultados explicam a alta variabilidade da POA, principalmente na primavera, como observado no painel direito da Figura 4.1. Além disso, a RMSP é afetada pela circulação da brisa marítima (Freitas et al., 2007; Vemado and Pereira Filho, 2016), uma fonte adicional de aerossol que também contribui para a alta variabilidade dos tipos de aerossol nessa região. Todas essas fontes de aerossol aumentam a variabilidade das propriedades do aerossol, tornando complexa a aquisição de dados de POA via satélite (Engel-Cox et al., 2006; Lee et al., 2016).

No sentido de especificar um modelo de aerossol que reflète as propriedades da camada de aerossol da RMSP, a pesquisa desenvolvida por de Almeida Castanho (2005) apresenta uma ferramenta que, utilizando uma propriedade dos aerossóis chamada refletância crítica, atribui dinamicamente um valor de AS ao modelo de aerossol, ao invés de utilizar valores de AS pré-fixados. No entanto, no que se refere às relações espectrais da superfície, ao contrário do algoritmo MAIAC, a metodologia desenvolvida por de Almeida Castanho (2005) utiliza uma razão pré-fixada de 0,6. Dessa forma, nossa sugestão para pesquisas futuras nessa área seria o desenvolvimento de um algoritmo que combine as melhorias propostas pelos algoritmos MAIAC e o proposto por de Almeida Castanho (2005): as especificações dinâmicas tanto das razões espectrais de superfície quanto do modelo de aerossol.

Além da POA obtida do MODIS-MAIAC, existem outros produtos de POA obtidos a partir dos dados do MODIS. O algoritmo *Dark Target* entrega os dados de POA nas resoluções espaciais de $10 \times 10 \text{ km}^2$ e $3 \times 3 \text{ km}^2$ (Levy et al., 2013). Já o algoritmo *Deep Blue* (Hsu et al., 2013) e o algoritmo que combina os dois algoritmos anteriores (Levy et al., 2013), entregam dados de POA com uma resolução espacial de $10 \times 10 \text{ km}^2$. Para estudos de escala local (i.e., regiões pequenas), como o caso da região da RMSP, essas baixas resoluções espaciais limitam a aplicação dos dados de POA.

O uso de dados de POA com alta resolução espacial, como os fornecidos pela versão 6 do algoritmo MAIAC, beneficia as aplicações em estudos epidemiológicos e de qualidade do ar, principalmente nas áreas urbanas (Chudnovsky et al., 2014a). No entanto, pelos resultados encontrados nesta tese, não recomendamos o seu uso para aplicações na RMSP.

Além da utilização dos dados de POA como uma aproximação para as concentrações de $MP_{2.5}$ na superfície, a outra contribuição da *Metodologia de Harvard* para o problema que foi estudado na tese foi a utilização do Modelo Linear Misto (MLM, Laird and Ware (1982)), que permite contemplar a variação diária dos fatores que influenciam a relação das concentrações de $MP_{2.5}$ com as variáveis meteorológicas e de uso e ocupação do solo. Então, mesmo que a POA não seja preditora das concentrações de $MP_{2.5}$ na RMSP, ainda podemos utilizar o MLM para a predição das concentrações de $MP_{2.5}$ nessa região.

Pensando na dinâmica das emissões veiculares, principal responsável pela composição do $MP_{2.5}$ na RMSP, propusemos MLMs que além das variáveis meteorológicas e de uso e ocupação do solo, incluíssem a variável indicadora de feriado, as variáveis indicadoras dos dias da semana, bem como o efeito de interação entre elas. Também propusemos MLMs nos quais foram incluídos os termos senos e cossenos para capturar o comportamento sazonal das concentrações do $MP_{2.5}$. E além disso, propusemos modelos que consideram as duas fontes de correlação inerentes aos dados: a espacial e a temporal.

O número reduzido de estações de monitoramento é um fator limitante para a aplicação da *Metodologia de Harvard* na RMSP, ou em regiões localizadas em países em desenvolvimento (Kloog et al., 2012). Com o intuito de aumentar a cobertura espacial do $MP_{2.5}$ na RMSP, consideramos que a concentração de $MP_{2.5}$ na RMSP corresponde a uma fração da concentração do MP_{10} (razão $MP_{2.5}/MP_{10}$). Vale destacar que o monitoramento do $MP_{2.5}$ em território nacional passou a ser previsto apenas em 2018, enquanto no Estado de São Paulo está previsto desde 2013. Já o monitoramento de MP_{10} está previsto desde 1990 com a Resolução CONAMA nº 3/1990. Isso explica o número relativamente baixo de estações medindo $MP_{2.5}$ comparado ao número de estações medindo MP_{10} .

Sobre a razão $MP_{2.5}/MP_{10}$, apesar de os relatórios de qualidade do ar da CETESB do período 2013-2018 divulgarem que na RMSP ela é cerca de 0,6, a análise dos dados das estações manuais para o período 2000-2014 mostrou que o valor dessa razão vem diminuindo na RMSP, e que no período estudado a razão está, em média, mais próxima de 0,5 do que de 0,6 (Apêndice A). Os dados das estações manuais produzem estimativas mais precisas da razão $MP_{2.5}/MP_{10}$ uma vez que fornecem as concentrações das frações $\leq 2,5$ e $2,5-10$ do MP. Ao utilizar os dados das estações automáticas para estimar a razão $MP_{2.5}/MP_{10}$ na RMSP, obtivemos valores de razão maiores ou muito próximos a 1, pois nas estações que medem esses dois poluentes são utilizados dois instrumentos independentes na mensuração de cada um deles.

Existem limitações ao se utilizar um valor constante para a razão $MP_{2.5}/MP_{10}$ para aumentar a cobertura espacial das estações medindo $MP_{2.5}$ na RMSP. Ao assumir que $MP_{2.5} \sim 0,5MP_{10}$ estamos assumindo que as concentrações dos poluentes é constante no tempo e

no espaço, o que pode não ser razoável. No entanto, esse tipo de abordagem costuma ser utilizado, por exemplo, em modelos de predição de concentração de poluentes desenvolvidos pela OMS (WHO, 2016).

Utilizando os dados das estações de monitoramento manual, tentamos caracterizar as variabilidades temporal e espacial da razão $MP_{2.5}/MP_{10}$ na RMSP. No entanto, a disponibilidade dos dados das estações manuais limitou a obtenção das estimativas das razões sazonais apresentadas nas Tabelas A.1 e A.2 para o período completo do nosso estudo, o que nos forneceria uma melhor caracterização da variação temporal da razão $MP_{2.5}/MP_{10}$ na RMSP. Já o número reduzido de estações manuais limitou a obtenção de uma estimativa da razão $MP_{2.5}/MP_{10}$ de acordo com a representatividade espacial das estações, o que seria importante para a caracterização da variabilidade espacial da razão $MP_{2.5}/MP_{10}$ na RMSP. Com a desativação da estação manual Ibirapuera em 19/12/2014 (CETESB, 2016b), o número de estações manuais medindo $MP_{2.5}$ e MP_{10} na RMSP tornou-se mais limitado.

Diante das limitações citadas anteriormente, fica fácil entender o porquê de o número de estudos que tiveram como objetivo quantificar as concentrações de MP na RMSP ser bastante reduzido. Até onde é de nosso conhecimento, o único modelo para a predição das concentrações de MP_{10} para a Cidade de São Paulo era o desenvolvido por Habermann and Gouveia (2012). Utilizando um Modelo de Regressão Múltipla, dados de nove estações de monitoramento da CETESB, dados demográficos, dados viários e dados de uso do solo, a metodologia proposta por esses autores foi capaz de explicar 64% da variabilidade da média anual da concentração de MP_{10} no ano de 2007. No modelo desenvolvido por esses autores apenas a variável “média diária do tráfego de veículos leves das vias coletoras, arteriais e de trânsito rápido até 250 m” permaneceu significativa no modelo final de predição. Uma das limitações do modelo desenvolvido por Habermann and Gouveia (2012) também foi o número reduzido de estações medindo o MP_{10} .

Considerando que os modelos ajustados nesta tese realizam predições em escala diária (um valor predito por dia, seja ele média diurna ou média diária), os resultados que encontramos são bastante satisfatórios. O desempenho dos modelos ajustados com os dados diários foi melhor do que o desempenho dos modelos ajustados com os dados diurnos, com valores de R^2 em torno de 0,60 e 0,70 e REQMP em torno de $4,7 \mu g/m^3$ e $4,1 \mu g/m^3$, respectivamente.

As hipóteses de que as variáveis feriado, dia da semana e os termos seno e cosseno melhoram o poder de predição dos modelos mostraram-se verdadeiras, tanto para os dados diurnos quanto para os dados diários, independente da estrutura de correlação considerada, porém os incrementos no poder de predição foram inferiores a 0,2% quando comparados com os modelos que utilizaram as variáveis do conjunto E1.

A hipótese de que modelar as duas fontes de correlação (espacial e temporal), melhoraria o poder de predição dos modelos também mostrou-se verdadeira. Fixado o conjunto de variáveis (E1, E2 e E3), comparando o modelo do tipo A2 com o modelo do tipo A1, houve um aumento da ordem de 5% no poder de predição, tanto para os modelos ajustados com

dados diurnos quanto para os modelos ajustados com dados diários.

Os resultados dos modelos do tipo A2 mostraram que os fatores que influenciam a relação das concentrações de $MP_{2.5}$ com as variáveis meteorológicas e de uso e ocupação do solo, variam mais ao longo dos dias, entre estações de monitoramento, do que intra-estação de monitoramento. Isso fica evidenciado ao analisarmos as razões entre as variâncias estimadas $\hat{\sigma}_u^2$ e $\hat{\sigma}_s^2$ dos efeitos aleatórios u_t e s_i , respectivamente, nos modelos do tipo A2 (Tabelas 4.5 e 4.6). Nos modelos ajustados com os dados diurnos e diários, essa razão foi em média igual a 6 e 9 respectivamente. A correlação temporal, introduzida nos modelos do tipo A2 por meio da inclusão dos efeitos aleatórios s_i , apesar de pequena, foi estatisticamente significativa e não pode ser ignorada, como na *Metodologia de Harvard*.

Apesar de as estimativas pontuais das concentrações de $MP_{2.5}$ utilizarem apenas as predições dos efeitos aleatórios u_t , nas duas estruturas de correlação consideradas, as estimativas obtidas a partir dos modelos que levaram em conta as duas fontes de correlação foram um pouco mais precisas, pois apresentaram, em média, menores valores de desvios-padrão dos erros de predição, como evidenciado nas Figuras 4.8 - 4.11.

Mesmo não utilizando os dados de POA, os resultados mostraram ainda que o desempenho dos modelos para as médias das concentrações diárias de $MP_{2.5}$ propostos na tese é similar ao desempenho dos modelos desenvolvidos para a Cidade do México, Suíça e Israel (Tabela 5.1). Os modelos desenvolvidos para essas regiões, além de variáveis meteorológicas e de uso e ocupação do solo, utilizaram dados de POA do algoritmo MAIAC e apresentaram valores de R^2 iguais a 0,72, 0,73 e 0,72, respectivamente, na primeira etapa da *Metodologia de Harvard*, enquanto os modelos desenvolvidos para as concentrações diárias de $MP_{2.5}$ nesta tese apresentaram valores de R^2 em torno de 0,70.

Tabela 5.1: Características dos estudos que aplicaram a Metodologia de Harvard

Referência	Área de estudo	Período do estudo	MP	Número de estações medindo MP	Resolução da POA	Satélite	R_{CV}^2
Lee et al. (2011)	Nordeste dos EUA	2003	$MP_{2.5}$	26	10 km^2	Terra e Aqua	0.92
Kloog et al. (2011)	Nordeste dos EUA	2000-2008	$MP_{2.5}$	78	10 km^2	Terra	0.83
Kloog et al. (2012)	Atlântico Médio dos EUA	2000-2008	$MP_{2.5}$	161	10 km^2	Terra	0.81
Chudnovsky et al. (2014b)	Nordeste dos EUA	2003	$MP_{2.5}$	62	1 km^2	Aqua	0.88
Lee et al. (2016)	Sudeste dos EUA	2003-2011	$MP_{2.5}$	257	10 km^2 e 1 km^2	Terra	0.73/ 0.76
Just et al. (2015)	Cidade do México	2004-2014	$MP_{2.5}$	12	1 km^2	Aqua	0.72
Kloog et al. (2015)	Israel	2003-2013	MP_{10} / $MP_{2.5}$	41/45	1 km^2	Aqua	0.79/ 0.72
Stafoggia et al. (2017)	Itália	2006-2012	MP_{10}	686	1 km^2	Aqua	0.65
Xiao et al. (2017)	China	2013-2014	$MP_{2.5}$	204	1 km^2	Terra e Aqua	0.76
de Hoogh et al. (2018)	Suíça	2003-2013	$MP_{2.5}$	99	1 km^2	Aqua	0.73

No que diz respeito às variáveis explicativas incluídas no modelo, a densidade viária, que foi utilizada como substituta das emissões veiculares, fator importante para a predição de $MP_{2.5}$ em áreas urbanas, foi estatisticamente significativa nos modelos do tipo A1 ajustados com dados diurnos e diários. Nos modelos do tipo A2 ajustados com dados diurnos e diários, a densidade viária apresentou níveis descritivos com valores em torno de 0,21 e 0,6,

respectivamente, não sendo portanto, uma variável significativa quando considerado o nível de significância de 5%.

Nos estudos de [Just et al. \(2015\)](#) e [de Hoogh et al. \(2018\)](#), na Cidade do México e na Suíça, respectivamente, a densidade viária foi um fator significativo nos modelos ajustados para explicar a variabilidade diária do $MP_{2.5}$, no entanto, a magnitude do efeito pode ser considerada marginal. O coeficiente associado à densidade viária foi igual a 0,21 no estudo de [Just et al. \(2015\)](#) e em média igual a 0,75 no estudo de [de Hoogh et al. \(2018\)](#), que ajustaram um modelo para cada ano do período compreendido entre 2003 e 2013, porém a densidade viária só foi significativa nos quatro primeiros anos daquele período. Nos modelos do tipo A1, os valores do coeficiente associado à densidade viária foram aproximadamente 3 e 1 nos modelos ajustados com dados diurnos e diários, respectivamente (Tabelas 4.5 e 4.6).

A magnitude da associação da densidade viária com as concentrações diurnas é maior do que a com as concentrações diárias de $MP_{2.5}$, pois é no período diurno (6-18h) que a evolução do perfil das emissões veiculares é mais parecida com a evolução do perfil das concentrações. Durante o período noturno observa-se uma evolução inversa dos perfis de emissão e de concentração (Figura 5.2). Apesar de as emissões veiculares serem menores durante a noite, o aumento da UR e a baixa ACLP nesse período favorecem o aumento das concentrações. Além disso, na RMSP existem outras emissões que geralmente não são consideradas nos inventários, como a emissão por pizzarias ([Mota Lima et al., 2020](#)).

Figura 5.2: Média das emissões de MP total (painel esquerdo) e médias das concentrações de $MP_{2.5}$ (painel direito) de acordo com a hora do dia. Os dados das emissões são do Inventário do IEMA e os dados das concentrações são das estações de monitoramento da CETESB, localizadas na Cidade de São Paulo, no ano de 2015. As linhas tracejadas verticais indicam o período diurno considerado na presente tese (6-18h). As linhas horizontais indicam as médias das concentrações/emissões nos diferentes intervalos de tempo (diurno, diário e noturno)

Na ausência da densidade viária, os modelos ficam dependentes das variáveis meteoro-

lógicas, e nesse caso, o método de interpolação utilizado na espacialização dessas variáveis pode interferir nos resultados obtidos. Foi o caso dos modelos do tipo A2 ajustados com dados diurnos e diários (Figuras 4.9 e 4.11). O padrão espacial esperado para as concentrações de $MP_{2.5}$ seria maiores concentrações no centro e menores concentrações na periferia da cidade. No entanto, na região da Serra da Cantareira (Zona Norte) observamos concentrações tão altas quanto as observadas no centro da cidade. Devido à ausência de estações meteorológicas nessa região (Figura 5.3) e devido ao tipo de interpolação utilizada (Método do Vizinho Mais Próximo), o perfil das variáveis meteorológicas nessa região acabou sendo representado pelo perfil mais urbano das estações Pirituba, Freguesia do Ó e Tremembé.

Uma opção seria utilizar algum método de interpolação que considerasse a relação já conhecida entre as variáveis meteorológicas e a elevação (Daly, 2006), como o método de interpolação polinomial. Considerando o caso da variável temperatura, apesar de a interpolação polinomial conseguir capturar as menores temperaturas esperadas na região da Serra da Cantareira, ela não captura as menores temperaturas também esperadas na Zona Sul da cidade, além de gerar um mapa em que quase não existe variação de temperatura nas demais regiões (Figura 5.3, painel esquerdo). Esse comportamento já era esperado, pois a RMSP está localizada em uma região com pouca variação de elevação (Planalto Atlântico). Por esse motivo, optamos por utilizar o método de interpolação do vizinho mais próximo.

Figura 5.3: Espacialização da variável temperatura no dia 27/02/2017, utilizando o método de interpolação polinomial ($Temperatura = a + b \times Elevacao$) (painel esquerdo) e o método do vizinho mais próximo (painel direito).

Modelos físico-químicos, como o CMAQ (*Community Multi-Scale Air Quality*, <https://www.epa.gov/cmaq>) e o WRF-Chem (*Weather Research e Forecasting with Chemistry*, Grell et al. (2005)) também podem ser utilizados para obter estimativas das concentrações de $MP_{2.5}$ em uma região. Esses modelos já foram utilizados no Brasil, mas para períodos de tempo muito restritos, geralmente, menores do que 1 mês (Andrade et al., 2015; Albuquerque et al., 2018; Vara-Vela et al., 2018; Albuquerque et al., 2019; Pedruzzi et al., 2019). Como são mais complexos, pois tentam reproduzir os processos de química atmosférica, de

transporte, de emissões e deposição dos poluentes, dependendo do tamanho do domínio, da resolução espacial empregada (regional ou urbana/local) e do nível de detalhamento das equações dos processos, geram uma maior demanda computacional, o que torna custosa a sua aplicabilidade para longos períodos de tempo, como o desta tese. Além de demandarem informações detalhadas sobre as emissões veiculares, industriais e naturais, que nem sempre estão disponíveis (Monn, 2001; Millar et al., 2010b; Pedruzzi et al., 2019).

No Brasil, Andreão et al. (2020) utilizaram o modelo *WRF-Chem* para as quantificar as concentrações anuais e diárias de MP_{10} e $MP_{2.5}$ em 102 cidades nas regiões metropolitanas da região sudeste no ano de 2015. Só para termos uma ideia do custo computacional, após a preparação de todos os arquivos de entrada (a parte mais trabalhosa de todo o processo) e utilizando 20 núcleos de processamento de um Intel®Xeon®, para cada região metropolitana a simulação das concentrações de um mês demorou cerca de 24h, considerando um modelo químico simplificado e a grade mais fina com resolução igual a 5 km^2 (comunicação via e-mail com Willian Andreão realizada em 2020). O ajuste dos modelos desta tese não demorou nem um minuto em um notebook com 16 GB de memória RAM e processador Intel®Core™i7. A parte mais demorada foi o cálculo dos desvios-padrão dos erros de predição que demorou no máximo 5h para um dos modelos.

Na Tabela 5.2 apresentamos a comparação entre o desempenho do *WRF-Chem* ajustado por Andreão et al. (2020) e o desempenho dos modelos E1A1 e E1A2, mas considerando apenas as concentrações diárias de $MP_{2.5}$ em 2015. O desempenho dos modelos ajustados mostrou-se bastante superior. Talvez a comparação com o *WRF-Chem* não seja justa, pois a resolução da grade mais fina utilizada foi igual a 5 km^2 enquanto a dos modelos ajustados foi igual a 1 km^2 .

Tabela 5.2: *Correlação linear (R) entre os valores observados e preditos pelo WRF-Chem (Andreão et al., 2020) e pelos modelos de predição E1A1 e E1A2 ajustados na presente tese com dados de 2015 - concentrações diárias de $MP_{2.5}$.*

Estação	Andreão et al. (2020)	E1A1	E1A2
USP-IPEN	0,64	0,900	0,902
Congonhas	0,56	0,898	0,900
Guarulhos - Pimentas	0,47	0,892	0,889
Ibirapuera	0,57	0,789	0,793
Itaim Paulista	0,6	0,887	0,887
Marginal Tietê - Ponte dos Remédios	0,62	0,891	0,889
Parelheiros	0,72	0,880	0,878
Pinheiros	0,65	0,916	0,918
São Bernardo do Campo - Centro	0,49	0,896	0,898

De acordo com Hoek et al. (2008), que realizaram uma revisão dos estudos que utilizaram modelos de regressão baseada em variáveis de uso e ocupação do solo para avaliar a variação espacial de poluentes, esses modelos apresentam desempenho semelhante ou superior ao desempenho dos modelos físico-químicos. Até o fechamento da presente tese, desconhecíamos

algum estudo que comparou o desempenho dos MLMs, utilizado na *Metodologia de Harvard*, com o desempenho de modelos físico-químicos na predição das concentrações de $MP_{2.5}$.

Apesar de mais custosos do ponto de vista computacional, alguns estudos já vêm empregando modelos físico-químicos para a predição de $MP_{2.5}$ em longos períodos de tempo, visando as aplicações epidemiológicas/saúde (Jin et al., 2019). E também vem crescendo a aplicação das metodologias que combinam os modelos físicos-químicos com os modelos estatísticos (Beloconi et al., 2018; Knibbs et al., 2018; Meng et al., 2019; Van Donkelaar et al., 2019).

Sobre as dificuldades encontradas durante o desenvolvimento do trabalho, devido à limitações computacionais não foi possível realizar o cálculo do AIC condicional, como discutido na seção 3.3.3.1.2, nem a etapa de análise de resíduos dos modelos apresentados nas seções 4.1 e 4.3.

O ajuste de um modelo estatístico está atrelado à suposições que precisam ser verificadas. Tal verificação pode ser realizada por meio da análise gráfica dos resíduos do modelo. No entanto, a análise de resíduos em MLM é um pouco mais complexa do que a empregada em modelos lineares usuais, pois existem duas fontes de variação: a intra e a interunidades amostrais (Nobre and Singer, 2007; Singer et al., 2017).

Uma rotina computacional para análise de resíduos de MLM está disponível em www.ime.usp.br/~jmsinger, mas tivemos dificuldades em utilizá-la na presente tese devido aos tamanhos das matrizes envolvidas nos cálculos dos resíduos. Algumas operações, como inversão de matrizes, não conseguiram ser efetuadas por causa da quantidade de memória RAM disponível. Tentamos otimizar a rotina com o uso de pacotes como *bigmatrix*, *bigmemory* e *BigStat* do *software* R, mas não tivemos êxito. No contexto da tese, as matrizes envolvidas nos cálculos dos resíduos possuem grandes dimensões e operar com matrizes assim, além de um maior poder computacional, exige uma pesquisa mais detalhada em técnicas/cálculos de álgebra linear.

Outra limitação refere-se à modelagem da correlação serial existente entre as medidas de uma mesma estação de monitoramento. Como demonstrado na seção 4.3, a inclusão do efeito aleatório de estação, s_i , melhorou o poder de predição dos modelos do tipo A2 em cerca de 5% quando comparados aos modelos do tipo A1 (*Metodologia de Harvard*). Porém, nos modelos do tipo A2, estamos assumindo que essa correlação é constante ao longo do tempo, o que pode não ser razoável para dados de séries temporais. Ou seja, apesar de ser um avanço frente à estrutura de covariâncias proposta pela *Metodologia de Harvard*, a estrutura de covariâncias dos modelos do tipo A2 pode ser melhor especificada.

Os efeitos da má especificação da estrutura da covariâncias em MLM, mas no contexto de análise de dados longitudinais, foram estudados por LeBeau (2018), por exemplo. Por meio de estudos de simulação, o autor conclui que a má especificação da estrutura de covariâncias gera estimativas não-viesadas dos efeitos fixos, mas gera estimativas enviesadas dos erros-padrão desses efeitos, o que acaba aumentando o erro do tipo I. Além disso, gera predições enviesadas dos efeitos aleatórios. No contexto do problema estudado na tese ainda não

existem estudos que fizeram esse tipo de comparação.

No contexto dos MLMs, a correlação serial pode ser inserida na modelagem por meio da inclusão de efeitos aleatórios (como fizemos no presente estudo), ou modelando a estrutura de covariâncias dos erros aleatórios ϵ_{it} . A primeira opção foi preferida à segunda, pois tivemos dificuldades computacionais em modelar a estrutura de covariâncias dos ϵ_{it} . No *software* R, as duas funções comumente utilizadas para o ajuste de MLMs são a *lme()* do pacote *nlme* (Pinheiro et al., 2020) e a *lmer()* do pacote *lme4* (Bates et al., 2015). A função *lmer()*, além de apresentar melhor desempenho ao ajustar modelos com grandes conjuntos de dados, também é mais eficiente no ajuste de modelos com efeitos aleatórios cruzados ¹, ambas características encontradas nos dados analisados na presente tese. No entanto, a função *lmer()* não permite a atribuição de estruturas de covariâncias mais complexas, tanto para os efeitos aleatórios quanto para os erros aleatórios (Bates et al., 2015).

Apesar de a função *lme()* permitir a atribuição de estrutura de covariâncias mais complexas, tanto para os efeitos aleatórios quanto para os erros aleatórios, nossa tentativa de ajuste de um modelo supondo uma estrutura de covariâncias auto-regressiva para os erros aleatórios, demorou alguns dias para gerar um resultado que não convergiu para uma solução.

Recentemente, encontramos o pacote *mgcv* do *software* R que apresenta alternativas para o ajuste de MLMs com grandes bancos de dados (Clark, 2019). Em trabalhos futuros pretendemos testar essa função nos dados utilizados na tese.

Mesmo diante das limitações expostas, os modelos para a predição da concentração do MP_{2.5} na RMSP propostos neste estudo são um avanço frente ao modelo proposto por Habermann and Gouveia (2012), pois como sugerido por esses autores, além de uma variável relacionada ao tráfego de veículos, incluem variáveis meteorológicas. Inclusive, no caso dos modelos do tipo A2, as variáveis meteorológicas mostraram-se mais importantes do que a variável relacionada ao tráfego de veículos para explicar a variabilidade das concentrações de MP_{2.5}. Além disso, os modelos desenvolvidos na presente tese fornecem estimativas em uma escala temporal mais fina, o que do ponto de vista epidemiológico, permite o estudo de desfechos a curto e a longo prazo.

Outra sugestão apresentada por Habermann (2012) foi a de incluir na modelagem variáveis que representassem a variação temporal do tráfego veicular. No entanto, dada a dificuldade de se obter e manter inventários de emissão (veicular e/ou industrial) atualizados em São Paulo e em outros locais do mundo (Alonso et al., 2010; Gallardo et al., 2012), a utilização de dados de POA obtidos do MODIS-MAIAC poderia ter se mostrado como uma boa alternativa, já que apresentam boas resoluções espacial e temporal. No entanto, como demonstramos, não é sempre que esse tipo de dado pode ser utilizado.

¹Dizemos que um efeito A é cruzado com um efeito B quando todos os níveis de A são observados nos níveis de B. No contexto da tese, os níveis do fator “tempo” (dias) são observados nos níveis do fator “estação” (estações de monitoramento). Quando níveis específicos do fator A são observados nos níveis do fator B, dizemos que A é aninhado em B. Por exemplo, os alunos em diferentes turmas de uma escola. Apesar de todos serem alunos, os alunos da turma 1 são diferentes dos da turma 2, e assim por diante.

Os desafios da aquisição de dados de POA de sensoriamento remoto via satélites sobre a RMSP são grandes e mais pesquisas nessa área são necessárias. Mesmo não utilizando esses dados, que têm se mostrado como um fator importante na predição das concentrações de $MP_{2.5}$ na superfície em diferentes locais no mundo, a metodologia apresentada mostrou resultados bastante satisfatórios. As estimativas de $MP_{2.5}$ obtidas dos modelos desenvolvidos na presente tese podem ser utilizadas como uma aproximação para a exposição à poluição atmosférica dos habitantes da RMSP, em estudos epidemiológicos futuros. Porém, dada a atual distribuição espacial, tanto das estações de monitoramento quanto das estações meteorológicas, acreditamos que as estimativas das concentrações de $MP_{2.5}$ obtidas a partir dos modelos apresentados nesta tese são mais representativas da Cidade de São Paulo do que da RMSP.

6. Conclusões

Os modelos de predição apresentados nesta tese foram capazes de estimar de forma satisfatória as concentrações de $MP_{2.5}$, em uma escala local sobre a RMSP, no período de janeiro de 2012 a dezembro de 2017.

Encontramos resultados inesperados ao tentar calibrar para a RMSP o MLM proposto no primeiro estágio da *Metodologia de Harvard*. Ao contrário do que foi observado em outras regiões onde a *Metodologia de Harvard* já foi aplicada, a POA apresentou sinal negativo na metade dos modelos ajustados com os dados coletados para a RMSP no período estudado.

A POA com alta resolução espacial do algoritmo MAIAC (versão 6), utilizada pela primeira vez na RMSP, não foi preditora das concentrações de $MP_{2.5}$ medidas na superfície nessa região no período estudado, uma vez que o modelo de aerosol empregado por esse algoritmo não reflete as propriedades do aerosol dessa região.

Os modelos propostos, que não utilizaram os dados de POA do MODIS-MAIAC, apresentaram desempenho semelhante ao dos modelos ajustados em outras regiões onde a *Metodologia de Harvard* já foi utilizada. Os modelos propostos foram capazes de explicar cerca de 60% e 70% da variabilidade das concentrações diurnas e diárias de $MP_{2.5}$, respectivamente, na RMSP, no período 2012-2017.

Os modelos propostos que consideraram as duas fontes de correlação inerentes aos dados (espacial e temporal) apresentaram poder de predição cerca de 5% maiores quando comparado ao dos modelos que consideram apenas a fonte de correlação espacial, tanto para os dados diurnos quanto para os dados diários.

Independente da estrutura de correlação considerada, os modelos com as variáveis indicadoras de feriado e dos dias da semana, e em alguns casos o efeito de interação entre elas, apresentaram poder de predição cerca de 0,01% e 0,04% maiores, para os dados diurnos diários respectivamente, quando comparado ao dos modelos ajustados apenas com as variáveis meteorológicas e de uso e ocupação do solo.

De forma similar, tanto para os dados diurnos quanto para os dados diários, os modelos que incluíram os termos específicos com o intuito de capturar a sazonalidade das concentrações de $MP_{2.5}$ ao longo do tempo, apresentaram poder de predição cerca de 0,02% maiores do que o dos modelos ajustados apenas com as variáveis meteorológicas e de uso e ocupação do solo.

Além das predições pontuais das concentrações de $MP_{2.5}$, apresentamos uma medida de incerteza associada a essas predições: os desvios-padrão dos erros de predição. Os modelos de predição propostos que consideraram as duas fontes de correlação inerente aos dados foram os que apresentaram estimativas mais precisas.

APÊNDICES

A. Estudo da razão $MP_{2.5}/MP_{10}$ na RMSP no período 2000-2014

Para estudar o comportamento da razão $MP_{2.5}/MP_{10}$ na RMSP, utilizamos os dados das concentrações de $MP_{2.5}$ e MP_{10} das estações de monitoramento manual Cerqueira César, Pinheiros, Ibirapuera e São Caetano do Sul no período de 2000-2014, obtidas por meio do método de amostragem dicotômico. Desse modo, temos dados referentes às concentrações das frações 2.5 e 2.5-10 do MP, sendo que a concentração de MP_{10} é estimada a partir da soma das duas frações anteriores. As amostragens foram realizadas a cada seis dias por períodos contínuos de 24 horas (CETESB, 2008).

A Figura A.1 apresenta os gráficos *boxplot* da razão $MP_{2.5}/MP_{10}$ ao longo do período 2000-2014, enquanto as Tabelas A.1 e A.2 apresentam as medidas descritivas da razão $MP_{2.5}/MP_{10}$ nas estações de monitoramento manual da RMSP de acordo com as estações do ano. Observa-se uma tendência de queda nos valores da razão $MP_{2.5}/MP_{10}$ na RMSP no período estudado.

Figura A.1: Gráfico boxplot da razão $MP_{2.5}/MP_{10}$ em 2000-2014 na RMSP. As linhas laranja e vermelha, representam o valor da razão $MP_{2.5}/MP_{10}$ divulgado pela CETESB nos relatórios de qualidade do ar (60%) e o valor da $MP_{2.5}/MP_{10}$ utilizado no presente estudo (50%), respectivamente.

Tabela A.1: *Medidas descritivas da razão $MP_{2.5}/MP_{10}$ nas estações de monitoramento manual da RMSP em 2000-2007.*

Período	n	Total		n	Cerqueira César		n	Pinheiros*		n	Ibirapuera**		n	São Caetano do Sul	
		média (dp)	mediana (mín-máx)		média (dp)	mediana (mín-máx)		média (dp)	mediana (mín-máx)		média (dp)	mediana (mín-máx)		média (dp)	mediana (mín-máx)
2000	36	61,01 (9,20)	60,94 (37,50 - 84,00)	36	61,01 (9,20)	60,94 (37,50 - 84,00)									
DJF	12	61,54 (8,61)	62,53 (44,44 - 76,92)	12	61,54 (8,61)	62,53 (44,44 - 76,92)									
MAM	16	61,36 (9,32)	60,94 (43,75 - 84,00)	16	61,36 (9,32)	60,94 (43,75 - 84,00)									
JJA	5	57,04 (12,84)	59,02 (37,50 - 73,44)	5	57,04 (12,84)	59,02 (37,50 - 73,44)									
SON	3	63,60 (6,32)	63,89 (57,14 - 69,77)	3	63,60 (6,32)	63,89 (57,14 - 69,77)									
2001	192	58,35 (7,97)	58,82 (39,13 - 88,24)	67	63,14 (8,02)	62,75 (43,75 - 88,24)	71	55,88 (7,47)	56,52 (39,13 - 75,00)				54	55,65 (5,55)	55,90 (42,86 - 66,67)
DJF	37	57,93 (7,24)	60,00 (42,42 - 80,77)	14	61,19 (8,09)	61,77 (43,75 - 80,77)	11	55,92 (7,40)	60,00 (42,42 - 63,64)				12	55,96 (4,77)	55,90 (45,00 - 63,16)
MAM	41	57,66 (8,93)	56,82 (43,48 - 88,24)	15	63,49 (10,65)	62,96 (50,00 - 88,24)	11	54,36 (5,84)	56,52 (43,48 - 62,71)				15	54,23 (5,69)	55,17 (44,83 - 65,08)
JJA	49	59,47 (8,37)	60,00 (42,31 - 86,00)	15	64,62 (8,69)	64,15 (50,00 - 86,00)	22	57,31 (8,39)	59,42 (42,31 - 75,00)				12	57,00 (4,78)	57,18 (49,41 - 65,00)
SON	65	58,18 (7,51)	58,14 (39,13 - 72,00)	23	63,11 (5,50)	63,16 (52,78 - 72,00)	27	55,32 (7,48)	55,32 (39,13 - 66,67)				15	55,74 (6,66)	56,92 (42,86 - 66,67)
2002	280	56,13 (8,06)	57,14 (20,83 - 76,74)	74	58,50 (7,25)	59,00 (35,29 - 74,07)	117	55,82 (6,90)	57,14 (35,71 - 69,23)	35	51,24 (10,64)	53,06 (20,83 - 76,74)	54	56,74 (8,29)	57,36 (28,57 - 73,33)
DJF	62	53,75 (8,93)	55,09 (28,57 - 69,23)	16	56,75 (9,78)	59,98 (35,29 - 68,00)	29	54,10 (7,76)	54,84 (37,50 - 69,23)	5	45,36 (6,92)	47,06 (37,50 - 53,33)	12	52,40 (9,67)	55,32 (28,57 - 62,50)
MAM	76	56,97 (6,90)	57,29 (39,13 - 74,07)	15	61,40 (6,49)	62,07 (50,00 - 74,07)	39	56,93 (5,68)	57,45 (44,90 - 67,57)	8	48,90 (8,70)	47,86 (39,13 - 60,98)	14	56,95 (5,64)	56,35 (48,28 - 67,50)
JJA	86	56,34 (8,22)	57,50 (20,83 - 76,74)	27	57,69 (5,32)	58,54 (47,89 - 70,00)	32	56,73 (6,17)	57,96 (40,98 - 69,09)	15	52,48 (13,13)	53,66 (20,83 - 76,74)	12	57,10 (10,12)	57,63 (30,00 - 69,81)
SON	56	57,31 (7,96)	58,01 (35,71 - 73,33)	16	58,87 (7,69)	60,15 (44,44 - 71,74)	17	54,50 (8,84)	56,25 (35,71 - 68,42)	7	55,44 (7,63)	53,85 (45,45 - 70,59)	16	59,54 (6,93)	59,98 (50,00 - 73,33)
2003	195	54,42 (7,93)	55,26 (29,69 - 71,05)	52	58,09 (6,28)	58,48 (36,36 - 68,18)	34	53,51 (8,06)	55,28 (33,33 - 65,75)	52	49,76 (8,68)	50,51 (29,69 - 67,86)	57	55,87 (6,29)	55,56 (39,39 - 71,05)
DJF	56	53,86 (7,00)	54,51 (33,33 - 66,67)	15	58,10 (5,31)	58,06 (44,00 - 66,67)	9	52,67 (5,83)	52,38 (45,45 - 63,64)	16	49,79 (8,65)	51,09 (33,33 - 61,76)	16	54,61 (4,82)	55,17 (40,91 - 61,82)
MAM	55	52,40 (9,31)	55,00 (33,33 - 71,05)	13	55,94 (8,66)	57,78 (36,36 - 65,79)	14	49,55 (8,70)	53,50 (33,33 - 60,27)	14	48,83 (10,26)	48,36 (33,33 - 64,00)	14	55,54 (7,99)	57,25 (39,39 - 71,05)
JJA	50	57,59 (6,72)	57,95 (39,58 - 69,05)	14	59,17 (5,50)	60,38 (47,83 - 66,67)	11	59,25 (5,49)	59,55 (46,67 - 65,75)	11	51,34 (7,20)	51,02 (39,58 - 67,86)	14	59,61 (5,89)	59,12 (51,35 - 69,05)
SON	34	53,98 (7,49)	54,65 (29,69 - 68,18)	10	59,37 (5,08)	58,48 (52,78 - 68,18)				11	49,35 (8,88)	50,00 (29,69 - 60,00)	13	53,77 (5,02)	53,33 (46,34 - 65,45)
2004	165	57,87 (8,23)	58,06 (33,33 - 85,37)	49	59,33 (6,82)	58,62 (38,46 - 75,00)				60	54,23 (8,27)	55,30 (33,33 - 70,00)	56	60,50 (8,04)	61,11 (39,39 - 85,37)
DJF	38	57,69 (9,13)	59,18 (33,33 - 74,07)	11	61,14 (8,23)	60,87 (42,86 - 74,07)				14	54,36 (10,25)	56,52 (33,33 - 68,75)	13	58,36 (7,93)	61,54 (42,11 - 66,67)
MAM	45	58,95 (5,86)	58,62 (43,75 - 75,00)	13	60,66 (6,00)	60,00 (50,00 - 75,00)				16	56,65 (6,31)	57,60 (43,75 - 66,67)	16	59,87 (4,81)	60,56 (50,00 - 70,37)
JJA	40	59,91 (8,95)	58,30 (42,25 - 85,37)	13	59,40 (5,82)	57,63 (50,88 - 75,00)				15	55,85 (7,92)	56,67 (42,25 - 70,00)	12	65,53 (10,47)	63,30 (55,56 - 85,37)
SON	42	54,94 (8,27)	54,17 (36,00 - 71,43)	12	56,15 (7,00)	56,79 (38,46 - 64,29)				15	49,92 (7,46)	50,00 (36,00 - 66,67)	15	59,00 (7,76)	60,38 (39,39 - 71,43)
2005	118	57,27 (8,41)	58,48 (31,25 - 72,41)	51	57,56 (8,40)	59,46 (31,25 - 71,43)				10	54,10 (10,20)	53,75 (41,82 - 70,00)	57	57,56 (8,14)	58,62 (35,00 - 72,41)
DJF	33	56,11 (9,27)	55,56 (36,67 - 72,41)	13	56,47 (8,07)	55,56 (40,63 - 71,43)				8	54,98 (10,37)	53,75 (42,11 - 70,00)	12	56,48 (10,46)	55,28 (36,67 - 72,41)
MAM	30	58,52 (6,86)	59,42 (41,67 - 68,18)	12	59,28 (6,37)	59,96 (44,44 - 68,18)				2	50,60 (12,41)	50,60 (41,82 - 59,38)	16	58,95 (6,47)	59,31 (41,67 - 67,86)
JJA	28	57,39 (6,91)	59,14 (36,67 - 70,97)	13	58,52 (7,13)	60,34 (44,83 - 70,97)							15	56,41 (6,80)	59,09 (36,67 - 65,12)
SON	27	57,16 (10,36)	60,00 (31,25 - 70,00)	13	56,13 (11,54)	59,46 (31,25 - 70,00)							14	58,12 (9,47)	60,49 (35,00 - 68,42)
2006	210	54,16 (8,66)	55,14 (24,32 - 77,78)	54	57,64 (7,55)	58,58 (41,18 - 77,59)	52	52,09 (7,37)	52,26 (36,36 - 66,67)	48	50,60 (10,87)	50,00 (24,32 - 77,78)	56	55,77 (6,98)	56,70 (40,00 - 69,23)
DJF	52	55,00 (8,29)	55,09 (37,50 - 77,78)	14	57,72 (6,02)	57,27 (50,00 - 72,73)	11	53,48 (7,54)	52,17 (42,86 - 65,22)	13	52,31 (11,33)	55,00 (37,50 - 77,78)	14	55,98 (7,27)	54,92 (43,64 - 69,23)
MAM	62	51,82 (8,69)	53,39 (24,32 - 77,59)	15	55,76 (9,39)	58,00 (41,18 - 77,59)	16	49,19 (6,70)	49,81 (36,36 - 58,82)	15	48,35 (10,49)	47,22 (24,32 - 66,67)	16	54,01 (6,08)	55,22 (41,67 - 62,32)
JJA	50	55,00 (8,43)	57,49 (34,78 - 66,15)	12	58,25 (7,01)	60,35 (45,65 - 66,15)	12	53,15 (8,11)	55,35 (40,00 - 61,11)	13	51,55 (10,51)	52,00 (34,78 - 66,07)	13	57,16 (6,50)	59,46 (41,30 - 64,18)
SON	46	55,43 (8,92)	57,14 (29,17 - 71,43)	13	59,17 (7,56)	60,00 (45,00 - 68,75)	13	53,49 (7,14)	52,34 (43,64 - 66,67)	7	50,46 (13,05)	50,00 (29,17 - 71,43)	13	56,33 (8,42)	60,00 (40,00 - 67,74)
2007	192	53,90 (8,13)	54,49 (29,73 - 73,68)	50	56,61 (7,00)	57,14 (34,43 - 73,08)	53	52,02 (8,03)	53,49 (30,95 - 72,00)	55	51,11 (8,85)	50,00 (29,73 - 73,68)	34	57,37 (6,17)	56,84 (43,75 - 72,00)
DJF	48	53,48 (8,28)	54,73 (31,58 - 68,18)	13	55,01 (8,72)	57,58 (34,43 - 68,18)	10	51,65 (9,45)	54,13 (31,58 - 65,00)	15	51,45 (9,25)	50,00 (33,33 - 66,67)	10	56,38 (3,28)	56,00 (52,00 - 62,50)
MAM	55	55,57 (7,29)	56,52 (39,39 - 69,23)	12	57,68 (6,13)	58,12 (50,00 - 66,67)	14	53,38 (6,47)	54,29 (39,39 - 61,90)	14	53,32 (8,70)	51,19 (40,91 - 66,67)	15	58,01 (6,79)	56,86 (43,75 - 69,23)
JJA	50	56,18 (7,92)	55,85 (36,17 - 73,68)	14	58,18 (7,90)	57,75 (44,83 - 73,08)	15	56,20 (5,92)	55,70 (46,15 - 72,00)	12	52,90 (9,87)	52,85 (36,17 - 73,68)	9	57,40 (7,90)	57,32 (46,67 - 72,00)
SON	39	49,16 (7,57)	50,00 (29,73 - 61,54)	11	55,35 (4,02)	55,56 (48,28 - 61,54)	14	46,45 (7,79)	47,17 (30,95 - 60,00)	14	47,01 (6,99)	47,49 (29,73 - 58,82)			

*: estação sem representatividade nos anos de 2004 e 2005; **: estação sem representatividade nos anos de 2002 e 2005;

dp: desvio padrão; mín: valor mínimo; máx: valor máximo; DJF: Dezembro, Janeiro, Fevereiro; MAM: Março, Abril, Maio; JJA: Junho, Julho, Agosto; SON: Setembro, Outubro, Novembro.

Tabela A.2: *Medidas descritivas da razão $MP_{2.5}/MP_{10}$ nas estações de monitoramento manual da RMSP em 2008-2014.*

Período	n	Total		n	Cerqueira César		n	Pinheiros		n	Ibirapuera		n	São Caetano do Sul	
		média (dp)	mediana (mín-máx)		média (dp)	mediana (mín-máx)		média (dp)	mediana (mín-máx)		média (dp)	mediana (mín-máx)		média (dp)	mediana (mín-máx)
2008	192	52,77 (8,58)	52,28 (23,81 - 81,82)	52	57,73 (8,08)	57,60 (37,50 - 81,82)	59	48,59 (6,95)	47,83 (35,00 - 64,71)	57	51,96 (9,24)	52,17 (23,81 - 81,25)	24	54,21 (5,71)	53,94 (40,00 - 62,50)
DJF	42	54,68 (8,45)	54,88 (33,33 - 70,00)	10	61,92 (5,42)	62,59 (52,94 - 70,00)	14	52,00 (8,31)	50,00 (42,86 - 64,71)	14	52,18 (8,55)	51,79 (33,33 - 66,67)	4	54,71 (5,96)	55,74 (47,37 - 60,00)
MAM	47	50,66 (7,10)	50,00 (35,00 - 64,71)	15	55,19 (6,70)	55,56 (43,48 - 64,71)	16	47,74 (6,47)	50,00 (35,00 - 61,54)	16	49,34 (6,29)	50,00 (37,50 - 57,14)			
JJA	46	52,98 (8,90)	52,56 (36,89 - 81,82)	13	58,12 (10,38)	55,88 (42,42 - 81,82)	14	48,61 (6,75)	47,51 (38,33 - 64,44)	14	52,46 (8,55)	54,17 (36,89 - 66,67)	5	53,33 (5,31)	52,94 (47,89 - 61,11)
SON	57	52,93 (9,32)	53,33 (23,81 - 81,25)	14	57,11 (8,14)	58,42 (37,50 - 70,00)	15	46,31 (5,55)	45,95 (35,71 - 57,14)	13	54,42 (13,20)	53,85 (23,81 - 81,25)	15	54,36 (6,12)	54,55 (40,00 - 62,50)
2009	223	53,65 (7,70)	53,57 (31,25 - 75,00)	48	57,25 (7,23)	57,14 (38,46 - 70,59)	56	50,92 (7,00)	51,20 (33,33 - 68,42)	59	52,68 (8,66)	52,94 (31,25 - 75,00)	60	54,29 (6,55)	53,92 (36,36 - 72,22)
DJF	51	52,13 (7,93)	52,94 (31,25 - 70,59)	12	55,51 (9,12)	56,51 (38,46 - 70,59)	12	51,14 (7,83)	51,74 (33,33 - 59,26)	13	49,38 (7,59)	51,85 (31,25 - 60,00)	14	52,65 (6,88)	52,98 (36,36 - 62,96)
MAM	59	51,36 (4,81)	51,85 (36,84 - 64,00)	14	52,99 (2,75)	52,59 (48,84 - 58,33)	15	49,01 (4,57)	47,83 (40,91 - 56,52)	15	50,07 (6,26)	51,61 (36,84 - 59,09)	15	53,46 (3,67)	53,49 (47,62 - 64,00)
JJA	56	55,86 (9,15)	57,14 (36,67 - 75,00)	10	59,81 (6,68)	61,16 (47,83 - 70,00)	14	53,36 (8,47)	54,07 (38,89 - 68,42)	16	55,50 (11,26)	57,14 (38,46 - 75,00)	16	55,92 (8,66)	56,83 (36,67 - 72,22)
SON	57	55,23 (7,59)	56,00 (33,33 - 70,59)	12	61,82 (6,32)	63,07 (45,45 - 70,59)	15	50,38 (6,84)	52,17 (33,33 - 58,82)	15	55,13 (7,30)	56,00 (44,44 - 66,67)	15	54,91 (6,04)	56,52 (46,15 - 64,00)
2010	224	52,58 (9,25)	52,36 (26,32 - 76,92)	56	58,61 (8,52)	60,36 (40,00 - 76,92)	58	50,54 (7,90)	50,00 (26,67 - 70,73)	50	48,19 (9,65)	48,19 (26,32 - 72,97)	60	52,60 (7,95)	52,14 (35,29 - 73,33)
DJF	48	52,06 (9,17)	52,09 (29,63 - 68,75)	11	60,20 (5,60)	61,54 (46,15 - 68,75)	14	48,90 (6,65)	48,14 (40,00 - 62,86)	9	44,94 (10,44)	41,18 (29,63 - 60,00)	14	53,39 (8,05)	52,62 (35,29 - 66,67)
MAM	59	54,23 (8,05)	53,85 (39,13 - 76,92)	16	61,49 (8,32)	61,73 (42,86 - 76,92)	15	51,89 (5,51)	52,63 (40,91 - 61,29)	12	49,84 (6,90)	49,39 (40,00 - 60,42)	16	52,44 (6,15)	52,45 (39,13 - 62,50)
JJA	59	51,66 (9,65)	50,60 (26,32 - 69,23)	15	58,55 (8,95)	60,00 (40,00 - 69,23)	14	49,31 (8,35)	50,00 (26,67 - 61,11)	15	47,74 (10,62)	48,24 (26,32 - 66,67)	15	50,88 (7,39)	49,09 (38,89 - 67,33)
SON	58	52,29 (10,03)	50,00 (30,00 - 73,91)	14	54,14 (9,10)	54,27 (40,00 - 73,91)	15	51,85 (10,49)	50,00 (30,00 - 70,73)	14	49,36 (10,51)	47,99 (33,33 - 72,97)	15	53,75 (10,29)	52,54 (40,00 - 73,33)
2011	217	51,84 (9,42)	52,00 (23,68 - 80,00)	54	54,56 (8,76)	53,85 (33,33 - 71,88)	52	51,89 (8,70)	50,00 (37,50 - 72,73)	52	47,95 (10,47)	50,00 (23,68 - 80,00)	59	52,72 (8,74)	52,38 (35,59 - 73,68)
DJF	51	50,03 (7,77)	51,52 (25,81 - 62,50)	11	53,51 (4,62)	52,94 (44,44 - 61,90)	12	49,36 (7,37)	47,27 (40,74 - 62,50)	14	47,45 (10,19)	52,38 (25,81 - 60,00)	14	50,46 (6,95)	50,00 (36,84 - 60,87)
MAM	56	52,42 (7,94)	52,63 (33,33 - 66,67)	15	54,85 (8,22)	55,56 (33,33 - 66,67)	13	52,53 (8,26)	52,00 (37,50 - 64,52)	13	49,50 (8,75)	50,00 (33,33 - 61,90)	15	52,43 (6,45)	52,38 (37,04 - 63,46)
JJA	56	54,82 (10,92)	54,65 (23,68 - 72,73)	15	57,78 (10,36)	63,04 (35,94 - 68,97)	13	56,81 (9,90)	55,36 (41,07 - 72,73)	12	47,95 (11,57)	51,45 (23,68 - 65,22)	16	55,58 (10,49)	54,65 (35,59 - 71,43)
SON	54	49,83 (9,94)	50,00 (29,17 - 80,00)	13	51,39 (9,59)	50,72 (36,36 - 71,88)	14	48,88 (7,58)	50,00 (38,46 - 65,52)	13	46,93 (12,26)	42,86 (29,17 - 80,00)	14	52,03 (10,23)	52,15 (36,36 - 73,68)
2012	210	50,95 (7,69)	50,00 (26,32 - 75,00)	50	54,97 (6,23)	54,92 (43,33 - 75,00)	56	48,87 (6,91)	48,20 (36,00 - 68,63)	50	48,21 (8,73)	49,14 (26,32 - 71,43)	54	51,94 (7,03)	52,33 (39,02 - 68,75)
DJF	47	51,83 (7,23)	50,00 (38,89 - 71,43)	11	53,73 (5,06)	54,84 (47,06 - 62,50)	11	49,50 (5,62)	47,37 (43,33 - 62,50)	13	50,25 (9,34)	50,00 (38,89 - 71,43)	12	53,92 (7,40)	54,09 (40,91 - 66,67)
MAM	53	49,35 (5,41)	50,00 (37,50 - 59,09)	13	52,94 (3,19)	52,94 (47,06 - 58,06)	15	46,81 (6,22)	45,45 (37,50 - 59,09)	11	49,56 (3,81)	50,00 (43,48 - 55,56)	14	48,58 (5,81)	48,42 (40,00 - 58,06)
JJA	52	52,43 (7,16)	52,47 (37,93 - 75,00)	13	57,07 (7,96)	55,77 (43,40 - 75,00)	14	50,42 (4,91)	49,39 (43,33 - 62,50)	11	47,29 (6,46)	48,28 (37,93 - 55,56)	14	54,16 (5,93)	53,50 (46,00 - 68,75)
SON	58	50,39 (9,84)	51,29 (26,32 - 69,49)	13	55,96 (7,22)	56,41 (43,33 - 69,49)	16	49,00 (9,50)	48,97 (36,00 - 68,63)	15	46,13 (11,89)	44,68 (26,32 - 66,67)	14	51,37 (8,04)	52,28 (39,02 - 66,10)
2013	181	49,95 (9,67)	50,00 (17,65 - 77,78)	52	52,55 (8,74)	53,24 (25,61 - 66,67)	17	49,82 (7,44)	50,00 (33,33 - 64,71)	56	46,21 (10,76)	46,05 (17,65 - 68,75)	56	51,31 (9,00)	51,50 (27,27 - 77,78)
DJF	42	49,00 (7,71)	50,00 (25,61 - 58,14)	10	50,21 (11,22)	54,34 (25,61 - 57,89)	9	49,69 (7,06)	50,00 (33,33 - 57,14)	12	46,35 (7,26)	50,00 (30,77 - 55,00)	11	50,22 (4,55)	50,00 (41,67 - 58,14)
MAM	52	47,10 (10,43)	47,28 (17,65 - 64,71)	14	50,94 (9,60)	51,32 (25,93 - 63,64)	8	49,97 (8,33)	46,25 (42,86 - 64,71)	15	40,53 (10,99)	38,89 (17,65 - 60,00)	15	48,56 (9,35)	50,00 (27,27 - 60,38)
JJA	47	53,07 (8,48)	52,08 (34,78 - 67,74)	15	55,93 (6,79)	55,26 (44,44 - 66,67)				16	49,23 (9,79)	48,10 (34,78 - 66,67)	16	54,22 (7,51)	53,55 (41,18 - 67,74)
SON	40	50,99 (10,84)	51,94 (24,24 - 77,78)	13	52,18 (7,40)	52,63 (41,03 - 66,67)				13	48,94 (12,68)	51,28 (24,24 - 68,75)	14	51,79 (12,13)	51,14 (32,35 - 77,78)
2014	159	48,62 (8,79)	50,00 (20,00 - 72,50)	48	52,35 (7,29)	52,12 (36,36 - 72,50)				56	44,29 (8,98)	44,44 (20,00 - 64,71)	55	49,76 (8,02)	50,00 (30,00 - 69,57)
DJF	36	49,28 (8,74)	49,53 (30,43 - 69,57)	10	51,92 (6,13)	49,24 (46,67 - 66,67)				13	44,19 (9,64)	42,86 (30,43 - 60,00)	13	52,33 (7,60)	50,00 (43,59 - 69,57)
MAM	41	45,92 (9,09)	48,00 (20,00 - 60,98)	13	49,20 (7,39)	50,00 (36,36 - 60,98)				13	41,83 (9,87)	47,37 (20,00 - 50,00)	15	46,62 (8,94)	50,00 (30,00 - 58,14)
JJA	38	50,23 (7,54)	50,00 (39,29 - 72,50)	11	54,64 (8,52)	53,13 (42,86 - 72,50)				15	47,22 (6,17)	44,68 (39,29 - 63,64)	12	49,94 (6,68)	49,32 (39,29 - 64,29)
SON	44	49,20 (9,26)	50,00 (28,57 - 68,97)	14	53,79 (6,58)	52,79 (39,02 - 67,65)				15	43,58 (10,01)	43,14 (28,57 - 64,71)	15	50,53 (8,14)	52,63 (35,29 - 68,97)

dp: desvio padrão; mín: valor mínimo; máx: valor máximo; DJF: Dezembro, Janeiro, Fevereiro; MAM: Março, Abril, Maio; JJA: Junho, Julho, Agosto; SON: Setembro, Outubro, Novembro.

B. *Script* para o ajuste dos modelos apresentados na seção 1.7

A seguir apresentamos os comandos utilizados para ajustar o modelo (1.1) apresentado na seção 1.7 e que os resultados foram apresentados na seção 4.1. Para exemplificar, estamos considerando os dados do período completo, as concentrações diárias de MP_{2.5} e os dados de POA do MODIS-MAIAC do satélite Terra.

```
#####  
## carrega as bibliotecas ##  
#####  
library(data.table)  
library(lme4)  
library(lmerTest)  
  
#####  
## para instalar pacotes: install.packages("nome do pacote") ##  
#####  
  
## define a localização dos dados  
dir_dados<-"~\\dados\\"  
  
## leitura dos dados ##  
load(paste0(dir_dados, "dados_tese_completos_diario.RData"))  
  
## os dados estão no objeto "dados" ##  
head(dados) ## exibe as primeiras linhas de "dados"  
  
##seleciona no banco de dados apenas as variáveis de interesse ##  
variaveis<-c("DATA","cell_number","estacao_nome", "estacao_id","nmp25","AOD_TERRA",  
"temperatura","umidade", "velocidade_vento", "precipitacao_acum", "HPBL_TERRA", "road_density")  
  
temp_data<-dados[,..variaveis]  
rm(dados)  
  
## cria variáveis para o ajuste do modelo ##  
  
## Altura da Camada Limite Planetária em km  
temp_data$HPBL_TERRA_km<-temp_data$HPBL_TERRA/1000  
  
## padronização da densidade viária pelo valor máximo  
temp_data$density<-temp_data$road_density/max(temp_data$road_density, na.rm=T)  
  
## cria a variável 'tempof' utilizada no ajuste do modelo  
temp_data$tempof<-factor(as.numeric(temp_data$DATA)-15339)  
  
## cria a variável 'estacao_id' utilizada no ajuste do modelo  
temp_data$estacao_id<-factor(temp_data$estacao_id)  
  
## ordena os dados de acordo com o 'id' da estação  
temp_data<-temp_data[order(temp_data$estacao_id),]  
  
## dados da etapa 1##  
## apenas as células com estação de monitoramento  
dados_etapa_1<-temp_data[!is.na(temp_data$estacao_nome), ]  
  
## ordena os dados pelo 'id' da estação e pela DATA  
dados_etapa_1<-dados_etapa_1[order(dados_etapa_1$estacao_id, dados_etapa_1$DATA),]  
  
## ajuste do modelo utilizando a função lmer do pacote lme4() ##  
fit<-lmerTest::lmer(nmp25~AOD_TERRA+temperatura+umidade+velocidade_vento+  
precipitacao_acum+HPBL_TERRA_km+density+ (AOD_TERRA|tempof),  
data=dados_etapa_1, na.action=na.omit, REML=F)  
summary(fit)
```


C. Equações e interpretação dos parâmetros dos modelos

Tabela C.1: *Equações dos modelos ajustados na tese.*

Modelo	Equação
E1A1	$MP_{it} = \alpha + \beta_1 Temp_{it} + \beta_2 UR_{it} + \beta_3 Prec_{it} + \beta_4 Vento_{it} + \beta_5 DViaria_i + u_t + \epsilon_{it}$
E2A1	$MP_{it} = \alpha + \beta_1 Temp_{it} + \beta_2 UR_{it} + \beta_3 Prec_{it} + \beta_4 Vento_{it} + \beta_5 DViaria_i + \beta_6 Feriado_t + \sum_{k=7}^{12} \beta_k I_{kt} + \sum_{k=13}^{18} \beta_k Feriado : I_{kt} + u_t + \epsilon_{it}$
E3A1	$MP_{it} = \alpha + \beta_1 Temp_{it} + \beta_2 UR_{it} + \beta_3 Prec_{it} + \beta_4 Vento_{it} + \beta_5 DViaria_i + \beta_6 Feriado_t + \sum_{k=7}^{12} \beta_k I_{kt} + \sum_{k=13}^{18} \beta_k Feriado : I_{kt} +$ $\sum_{j=1}^2 \left[\delta_j sen \left(\frac{2\pi jt}{365} \right) + \gamma_j \cos \left(\frac{2\pi jt}{365} \right) \right] + u_t + \epsilon_{it}$
E1A2	$MP_{it} = \alpha + \beta_1 Temp_{it} + \beta_2 UR_{it} + \beta_3 Prec_{it} + \beta_4 Vento_{it} + \beta_5 DViaria_i + u_t + s_i + \epsilon_{it}$
E2A2	$MP_{it} = \alpha + \beta_1 Temp_{it} + \beta_2 UR_{it} + \beta_3 Prec_{it} + \beta_4 Vento_{it} + \beta_5 DViaria_i + \beta_6 Feriado_t + \sum_{k=7}^{12} \beta_k I_{kt} + \sum_{k=13}^{18} \beta_k Feriado : I_{kt} + u_t + s_i + \epsilon_{it}$
E3A2	$MP_{it} = \alpha + \beta_1 Temp_{it} + \beta_2 UR_{it} + \beta_3 Prec_{it} + \beta_4 Vento_{it} + \beta_5 DViaria_i + \beta_6 Feriado_t + \sum_{k=7}^{12} \beta_k I_{kt} + \sum_{k=13}^{18} \beta_k Feriado : I_{kt} +$ $\sum_{j=1}^2 \left[\delta_j sen \left(\frac{2\pi jt}{365} \right) + \gamma_j \cos \left(\frac{2\pi jt}{365} \right) \right] + u_t + s_i + \epsilon_{it}$

Nos modelos definidos nas equações apresentadas na Tabela C.1, temos:

1. MP_{it} : média da concentração diária/diurna de $MP_{2.5}$ observada no dia t na estação de monitoramento i ;
2. α é o intercepto fixo enquanto u_t é o intercepto aleatório de dia (tempo);
3. s_i é o intercepto aleatório de estação;
4. $Temp_{it}$, $Vento_{it}$, UR_{it} e $Prec_{it}$ são os valores de temperatura da superfície, velocidade do vento, umidade relativa e precipitação acumulada, respectivamente, correspondentes à célula i no dia t ;
5. $DViaria_i$ é o valor calculado da densidade viária correspondente à célula i ;
6. $Feriado_t$ é a variável indicadora da ocorrência de feriado, assumindo o valor 1 se o dia t for feriado e o valor 0, caso contrário;
7. I_{kt} são variáveis indicadoras do dia da semana, assumindo valores 1 se o dia t corresponder ao dia da semana I_k (I_7 : segunda, I_8 : terça,..., I_{12} : sábado), sendo que o domingo foi utilizado como categoria de referência;
8. os termos $Feriado : I_{kt}$ representam a interação entre os dias da semana I_k e a ocorrência de feriado no dia t (Feriado:segunda, Feriado:terça,..., Feriado:sábado);
9. e os termos $\delta_j \sin\left(\frac{2\pi jt}{365}\right)$ e $\gamma_j \cos\left(\frac{2\pi jt}{365}\right)$ são responsáveis por capturar o comportamento sazonal da concentração de $MP_{2.5}$ ao longo do tempo;
10. nos modelos do tipo A1 (E1A1, E2A1 e E3A1) temos as componentes $u_t \sim N(0, \sigma_u^2)$ e $\epsilon_{it} \sim N(0, \sigma^2)$, mutuamente independentes.
11. e nos modelos do tipo A2 (E1A2, E2A2 e E3A2) temos as componentes $u_t \sim N(0, \sigma_u^2)$, $s_i \sim N[0, \sigma_s^2]$ e $\epsilon_{it} \sim N(0, \sigma^2)$, mutuamente independentes.

D. *Script* para o ajuste do modelo que considera apenas a correlação espacial

A seguir apresentamos os comandos utilizados para ajustar o modelo E1A1 apresentado na seção 4.3.1. Para exemplificar, estamos considerando as concentrações diárias de $MP_{2.5}$. Também apresentamos os comandos utilizados na etapa de validação cruzada, na predição das concentrações de $MP_{2.5}$, no cálculo dos desvios-padrão dos erros de predição e na confecção dos mapas de concentração.

```
#####  
## carrega as bibliotecas ##  
#####  
library(data.table)  
library(lme4)  
library(lmerTest)  
library(Matrix)  
library(gdata)  
library(raster)  
library(rgdal)  
library(lubridate)  
library(data.table)  
library(colorRamps)  
  
#####  
## definir localização dos dados e onde os resultados serão salvos ##  
#####  
dir_dados<-"~\\dados\\"  
dir_resultados<-"~\\resultados1\\"  
  
#####  
## leitura dos dados ##  
#####  
load(paste0(dir_dados, "dados_tese_completos_diario.RData"))  
  
#####  
## os dados estão no objeto "dados" ##  
#####  
head(dados) ## exibe as primeiras linhas de "dados"  
  
#####  
##seleciona apenas as variáveis de interesse ##  
#####  
variaveis<-c("DATA", "cell_number", "estacao_nome", "estacao_id", "nmp25", "temperatura", "umidade",  
"velocidade_vento", "precipitacao_acum", "road_density")  
  
temp_data<-dados[,..variaveis]  
rm(dados)  
  
#####  
## cria variáveis para o ajuste do modelo ##  
#####  
## cria a variável 'tempof'  
temp_data$tempof<-factor(as.numeric(temp_data$DATA)-15339)  
  
## cria a variável 'estacao_id' utilizada no ajuste do modelo  
temp_data$estacao_id<-factor(temp_data$estacao_id)  
  
## padroniza a variável denisdade viária pelo valor máximo observado  
temp_data$density<-temp_data$road_density/max(temp_data$road_density,na.rm=T)  
  
## ordena os dados de acordo com o 'id' da estação  
temp_data<-temp_data[order(temp_data$estacao_id),]
```

```
#####
### separa os dados da etapa 1 ###
#####

### apenas as células com estação de monitoramento
dados_etapa_1<-temp_data[!is.na(temp_data$estacao_nome), ]
### ordena os dados pelo 'id' da estação e pela DATA
dados_etapa_1<-dados_etapa_1[order(dados_etapa_1$estacao_id, dados_etapa_1$DATA),]

#####
### ajuste do modelo ElA1 (Tabela 4.6) utilizando a função lmer do pacote lme4() ###
#####
fit<-lmerTest::lmer(nmp25~temperatura+umidade+velocidade_vento+density+
(1|tempof),data=dados_etapa_1, na.action=na.omit, REML=F)
summary(fit)

#####
### Etapa de validação cruzada e cálculo das medidas de desempenho ###
#####

### função para a etapa de validação cruzada ###
funcao_CV<-function(s,dataset){
  dataset$id<-1:dim(dataset)[1]
  print(paste0("validação estação - ", s))
  dados_preditos_cv<-NULL
  cross_ajuste<-subset(dataset, cell_number!=s) ###dados ajuste
  cross_teste<-subset(dataset, cell_number==s) ###dados teste

  ### ajuste do modelo desconsiderando os dados da estação s ###
  fit_ajuste<-lmer(fit, data=cross_ajuste, na.action=na.omit)

  ### parte aleatória do modelo ###
  ### independente da estrutura de variâncias-covariâncias escolhida,
  ### só conhecemos o efeito aleatório do tempo
  b<-ranef(fit_ajuste)$tempof[[1]]

  ### construção da matriz Zi ###
  ### tempos não observados na estação s ###
  no_obs<-as.numeric(colnames(Z)[1:length(b)][!(colnames(Z)[1:length(b)]%in%unique(cross_teste$tempof))])
  if(length(no_obs)==0){
    Zi<-Diagonal(n=length(b), x=1)
  }else{
    Zi<-Diagonal(n=length(b), x=1)
    Zi[, no_obs]<-0 ### zera os tempos não observados em uma determinada estação
  }

  aleathat<-Zi%*%b
  aleathat<-aleathat[-c(no_obs)] ### exclui os dias faltantes nessa estação

  ### construção da matriz Xi ###
  ### extrai as partes da formula do modelo
  f0<-gsub(' \\+ \\(1 \\| tempof\\)', "", formula(fit_ajuste))
  ### constroi a matriz Xi do modelo criando as dummies (intercepto)
  Xi<-model.matrix(formula(paste0("~", f0[3])), data=cross_teste)

  ### parte fixa do modelo ###
  betha<-summary(fit_ajuste)$coeff[,1]
  fixedhat<-Xi%*%t(betha)

  ### predição marginal ###
  yhat.m<-fixedhat

  ### predição condicional ###
  yhat.c<-fixedhat+aleathat

  ### armazena os dados preditos para a estação s ###
  dados_preditos_cv<-data.frame(cross_teste, yhat_c=as.matrix(yhat.c),
  yhat_m=as.matrix(yhat.m), aleat=as.matrix(aleathat))

  return(dados_preditos_cv)
}

#####
### Validação Cruzada ###
#####
```



```

data.set<-na.omit(dados_etapa_1)
rm(dados_etapa_1)

### Matrizes para o cálculo das predições e Validação Cruzada ###
Z<-getME(fit,"Z")

### Validação Cruzada - células com estação de monitoramento ###
mylist<-NULL
mylist<-lapply(unique(data.set$cell_number), dataset=data.set, funcao_CV)
dados_preditos<-NULL
dados_preditos<-do.call('rbind', mylist)
rm(dados_preditos)

#####
### Cálculo das medidas de desempenho ###
#####

### Gráficos 'observado x predito' e medidas de desempenho por estação ###
tab_temp<-NULL
for(k in unique(dados_preditos$estacao_id)){
  dados_plot<-dados_preditos[dados_preditos$estacao_id==k,]

  print(paste0('desempenho estação - ', k))
  png(file=paste0(dir_resultados, k, "_observado_predito.png"), width=6*680, height=3*680, res=300)
  par(mfrow=c(1,1), mar=c(4,5,4,0))
  plot(dados_plot$DATA, dados_plot$ncmp25, type="l",
        ylab=expression("Média diária da concentração de MP"[2.5]*" ("*mu*"g/m^3*")"), xlab=" ",
        main=gsub("_", " ", paste0(unique(dados_plot$estacao_nome))),
        las=2, ylim=c(-5, max(dados_preditos$ncmp25, na.rm=T)), cex.main=1.3, cex.axis=1.3, cex.lab=1.3)
  lines(dados_plot$DATA, dados_plot$yhat_c, type="l", col='yellow', ylab=" ", cex=2)
  lines(dados_plot$DATA, dados_plot$yhat_m, type="l", col='red', ylab=" ", cex=2, lwd=2)
  legend("topleft", c("observado", "predição marginal", "predição condicional"), lty=1,
        col=c('black','red','yellow'), bty='n', cex=1.3)
  dev.off()

  EQMPc_temp<-sqrt(mean((dados_plot$ncmp25-dados_plot$yhat_c)^2, na.rm=T))
  R2c_temp<-summary(lm(dados_plot$ncmp25~dados_plot$yhat_c))$adj.r.squared

  tab_temp<-rbind(tab_temp, cbind(data.frame(unique(dados_plot$estacao_nome)), R2c_temp, EQMPc_temp))
}
### tabela com as medidas de desempenho de cada estação de monitoramento
tab_temp

### Medidas de desempenho global ###
R2_c<-summary(lm(dados_preditos$ncmp25~dados_preditos$yhat_c))$adj.r.squared
REQMP.c<-sqrt(mean((dados_preditos$ncmp25-dados_preditos$yhat_c)^2, na.rm=T))
round(summary(fit)$AICtab,4)[1]

#####
### Etapa 2 com cálculo dos desvios-padrão dos erros de predição (DPEP) ###
#####

### função para o cálculo das predições e do DPEP para cada célula ###
funcao_ET2_erro<-function(s,dataset){
  print(paste0("predição célula - ", s))
  cross_teste<-subset(dataset, cell_number==s)

  ### Cálculo das predições e do DPEP ###

  #independente do modelo só conhecemos o efeito aleatório do tempo
  b<-ranef(fit)$tempof[[1]]

  ### construção da matriz Zi"
  ### tempos não observados na estação s ###
  no_obs<-colnames(Z)[1:length(b)][!(colnames(Z)[1:length(b)]%in%unique(cross_teste$tempof))]
  if(length(no_obs)==0){
    Zi<-Diagonal(n=length(b), x=1)
  }else{
    Zi<-Diagonal(n=length(b), x=1)
    Zi[, no_obs]<-0 ### zera os tempos não observados em uma determinada estação
  }

  ### parte aleatória do modelo
  aleathat<-Zi%*%b

```

```

### construção da matriz Xi ###
##extrai as partes da formula do modelo selecionado
f0<-gsub(' \\+ \\(1 \\| tempof\\)', "", formula(fit))
Xi<-model.matrix(formula(paste0("~", f0[3])), data=cross_teste)

### parte fixa do modelo ###
betha<-summary(fit)$coeff[,1]
fixedhat<-Xi%*%t(t(betha))

### predição marginal ###
yhat.m<-fixedhat

### predição condicional ###
yhat.c<-fixedhat+aleathat

### variâncias dos erros de predição marginal - as variâncias estão na diagonal ###
var.m<-Xi%*%vcov(fit)%*%t(Xi)+Zi%*%G[1:length(b),1:length(b)]%*%t(Zi) +
Diagonal(n=dim(Xi)[1], x=s2) + 2*(Xi%*%cov_beta_b%*%t(Zi))

### variâncias dos erros de predição condicional - as variâncias estão na diagonal ###
var.c<-Xi%*%vcov(fit)%*%t(Xi)+ Zi%*%var_b%*%t(Zi) +
Diagonal(n=dim(Xi)[1], x=s2) + 2*(Xi%*%cov_beta_b%*%t(Zi))

dados_preditos_ET2<-NULL

### armazena as estimativas e os DPEP
dados_preditos_ET2<-cross_teste[, c("yhat_c", "var_c", "yhat_m", "var_m"):=list(as.matrix(yhat.c),
as.matrix(sqrt(diag(var.c))), as.matrix(yhat.m),as.matrix(sqrt(diag(var.m))))]

return(dados_preditos_ET2)
}

#####
### Matrizes e quantidades utilizadas nas predições e no DPEP ###
#####
data.set<-na.omit(dados_etapa_1)
### número de tempos/dias
n_tempo<-ncol(table(drop.levels(as.factor(data.set$estacao_id)), drop.levels(as.factor(data.set$tempo))))
rm(data.set)

### matriz de Var-Cov do efeito associado ao tempo
rvcov_tempo<-VarCorr(fit)$tempof
### quando tiver apenas o intercepto ###
rvcov_tempo<-attr(rvcov_tempo, "stddev")*attr(rvcov_tempo, "correlation")*attr(rvcov_tempo, "stddev")

### Matriz G ###
G_tempo<-kronecker(diag(n_tempo),rvcov_tempo)
G<-Matrix(G_tempo, sparse=T)
rm(G_tempo)

### estimativa de sigma2 e sigma2b###
s2<-(attr(VarCorr(fit), "sc"))^2
s2b<-as.numeric(rvcov_tempo)

### Matrizes para o cálculo dos DPEP ###
X<-getME(fit,"X") ### ordenado por estacao_id e DATA
y<-getME(fit,"y")

#####
### Matrizes do Sistema de Equações de Henderson (A) ###
#####
m11<-(1/s2)*(t(X)%*%X)
m12<-(1/s2)*(t(X)%*%Z)
M1<-cbind(m11, m12)
rm(m11)

m21<-(t(m12))
m22<-(1/s2)*(t(Z)%*%Z) + Matrix::diag(1/s2b, n_tempo)
M2<-cbind(m21, m22)
rm(m12)
rm(m21)
rm(m22)

A<-rbind(M1, M2)

```

```

rm(M1)
rm(M2)
beta<-summary(fit)$coeff[,1]

### cálculo da variância - inversa da matriz do Sistema de Equações de Henderson (A)
invA<-solve(A)

### var(b)
var_b<-invA[(length(beta)+1):(length(beta)+length(ranef(fit)$tempof[[1]])),
(length(beta)+1):(length(beta)+length(ranef(fit)$tempof[[1]]))]

cov_beta_b<-invA[1:length(beta), (length(beta)+1):(length(beta)+length(ranef(fit)$tempof[[1]]))]
rm(A)
rm(invA)

### conjunto de dados completo: dados das etapas 1 e 2 ###
data.set<-temp_data[order(temp_data$cell_number),]
data.set<-data.set[which(!is.na(data.set$road_density)),]
data.set$id<-1:dim(data.set)[1]
rm(temp_data)

### Cálculo das predições e dos DPEP para cada célula do grid###
### demorou cerca de 2h em um notebook com 8GB de RAM e processador Intel i7.
mylist<-NULL
mylist<-lapply(unique(data.set$cell_number), dataset=data.set, funcao_ET2_erro)
predicoes_erro_final<-NULL
print("rbind...")
predicoes_erro_final<-do.call('rbind', mylist)

### salva as predições e os respectivos DPEP em um arquivo ###
#save(predicoes_erro_final, file=paste0(dir_resultados, "predicoes_erro_final.RData"))

#####
### Mapas com as predições e com os DPEP ###
#####

### shapefile da RMSP - WGS84 UTM 23S
rmspta<-readOGR(paste0(dir_dados, 'rmsp\\rmsp_utm.shp'))

### shapefile da Cidade de SP - WGS84 UTM 23S
cidadeta<-readOGR(paste0(dir_dados, 'cidade\\cidade_utm.shp'))

### grid espaço-temporal para as médias das predições - anual ###
chave_cell<-expand.grid(1:15092, ymd(seq(as.Date("01/01/2012", format="%d/%m/%Y"),
as.Date("31/12/2017", format="%d/%m/%Y"), by="days")))
colnames(chave_cell)<-c("cell_number", "DATA")
chave_cell<-data.table(chave_cell)
keycols <- c("cell_number","DATA")
setkeyv(chave_cell, keycols)

### grid espaço-temporal para as predições de acordo com as estações do ano ###
chave_cell_season<-expand.grid(1:15092, c("DJF","MAM","JJA","SON"))
colnames(chave_cell_season)<-c("cell_number", "season")
chave_cell_season<-data.table(chave_cell_season)
keycols <- c("cell_number","season")
setkeyv(chave_cell_season, keycols)

### esquema de cores dos mapas ###
### médias anual e sazonais
pal <- colorRampPalette(c("green","yellow","orange","red"))
### DPEP
pal_erro <- colorRampPalette(c("white","red","brown","black"))

### seleciona as colunas de interesse - predições e DPEP condicionais ###
### apesar de chamar 'var' já é o desvio padrão
dados_plot_temp<-predicoes_erro_final[, c("DATA","cell_number","yhat_c","var_c")]
rm(predicoes_erro_final)

### merge com a grade espaco-temporal ###
dados_plot_temp<-merge(chave_cell, dados_plot_temp, by=c("cell_number", "DATA"), all=T)

### cria as variáveis 'ano', 'mes' e 'season' (estacao do ano) ###
dados_plot_temp$ano<-year(dados_plot_temp$DATA)
dados_plot_temp$mes<-month(dados_plot_temp$DATA)
dados_plot_temp$season<-ifelse(dados_plot_temp$mes%in%c(12,1,2),"DJF",

```

```

        ifelse(dados_plot_temp$mes%in%c(3,4,5),"MAM",
              ifelse(dados_plot_temp$mes%in%c(6,7,8),"JJA","SON"))

### médias das predições no período completo (2012-2017) sobre a RMSP ###
dados_plot_temp2<-dados_plot_temp[,.(yhat_c=mean(yhat_c), var_c=mean(var_c)), by="cell_number"]

### médias das predições por estacao do ano ###
dados_plot_season<-dados_plot_temp[,.(yhat_c=mean(yhat_c), var_c=mean(var_c)), by=list(cell_number, season)]

dados_plot_season<-merge(chave_cell_season, dados_plot_season, by=c("cell_number", "season"), all=T)
rm(dados_plot_temp)

### transformando os valores preditos em raster - anual RMSP ###
est_RMSP <- raster::raster(extent(rmspta), res=1000)
proj4string(est_RMSP) <- TA
values(est_RMSP)<-dados_plot_temp2$yhat_c

### DPEP - anual RMSP ###
erro_RMSP <- raster::raster(extent(rmspta), res=1000)
proj4string(erro_RMSP) <- TA
values(erro_RMSP)<-dados_plot_temp2$var_c

### raster com as predições e os DPEP ###
raster_RMSP<-stack(est_RMSP, erro_RMSP)
names(raster_RMSP)<-c('yhat_c','var_c')
rm(est_RMSP)
rm(erro_RMSP)
rm(dados_plot_temp2)

### transformando os valores preditos em raster - estacoes do ano RMSP ###
est_DJF_RMSP <- raster::raster(extent(rmspta), res=1000)
proj4string(est_DJF_RMSP) <- TA
values(est_DJF_RMSP)<-dados_plot_season$yhat_c[dados_plot_season$season=="DJF"]

est_MAM_RMSP <- raster::raster(extent(rmspta), res=1000)
proj4string(est_MAM_RMSP) <- TA
values(est_MAM_RMSP)<-dados_plot_season$yhat_c[dados_plot_season$season=="MAM"]

est_JJA_RMSP <- raster::raster(extent(rmspta), res=1000)
proj4string(est_JJA_RMSP) <- TA
values(est_JJA_RMSP)<-dados_plot_season$yhat_c[dados_plot_season$season=="JJA"]

est_SON_RMSP <- raster::raster(extent(rmspta), res=1000)
proj4string(est_SON_RMSP) <- TA
values(est_SON_RMSP)<-dados_plot_season$yhat_c[dados_plot_season$season=="SON"]

### raster com as predições de todas as estações do ano ###
est_season_RMSP <- stack(est_DJF_RMSP,est_MAM_RMSP,est_JJA_RMSP,est_SON_RMSP)
names(est_season_RMSP)<-c("DJF", 'MAM', "JJA", "SON")
rm(est_DJF_RMSP)
rm(est_MAM_RMSP)
rm(est_JJA_RMSP)
rm(est_SON_RMSP)
rm(dados_plot_season)

#####
### Mapas para a RMSP e para a Cidade de SP ###
#####

### Como no ArcGis e no plot (default) do R, ajustamos os valores plotados
### de acordo com os percentis. Isso evita distorção na escala devido aos valores
### muito altos/pequenos ###

### vou ajustar pelos percentis: 5% e 95%
percentil=0.05 #opcões 0,0.005,0.01,0.025 ou 0.05

### Mapas: para RMSP use regiao="RMSP"
### para a Cidade de São Paulo use regiao="SP"
### e execute apenas os próximos comandos

regiao="RMSP"

if(regiao=="RMSP"){
  raster_1<-raster_RMSP
  raster_2<-est_season_RMSP

```

```

    contorno<-rmspta
  }else{
    ### estimativas anuais para a cidade de SP ###
    raster_SP<-crop(raster_RMSP,extent(cidadeta)) ### corta o extent
    raster_SP<-mask(raster_SP,cidadeta) ### corta o formato

    ### raster estimativas por estação do ano - cidade de SP ###
    est_season_SP<-crop(est_season_RMSP,extent(cidadeta)) ### corta o extent
    est_season_SP<-mask(est_season_SP,cidadeta) ### corta o formato
    raster_1<-raster_SP
    raster_2<-est_season_SP
    contorno<-cidadeta
  }

  ### cálculo dos percentis (nas duas caudas)
  breakpoints<-quantile(raster_1$yhat_c, c(0,0.005,0.01,0.025,0.05,0.95,0.975,0.99,0.995,1))
  raster_1_temp<-raster_1
  breakpoints_season<-quantile(c(values(raster_2)),
    c(0,0.005,0.01,0.025,0.05,0.95,0.975,0.99,0.995,1), na.rm=T)
  raster_2_temp<-raster_2
  rm(raster_1)
  rm(raster_2)

  ### tudo que for menor do que percentil recebe o valor do percentil
  ### tudo que for maior do que 1-percentil recebe o valor do 1-percentil
  if(percentil==0.01){
    values(raster_1_temp$yhat_c)[values(raster_1_temp$yhat_c)<=breakpoints[2]]<-breakpoints[2]
    values(raster_1_temp$yhat_c)[values(raster_1_temp$yhat_c)>=breakpoints[9]]<-breakpoints[9]

    values(raster_2_temp)[values(raster_2_temp)<=breakpoints_season[2]]<-breakpoints_season[2]
    values(raster_2_temp)[values(raster_2_temp)>=breakpoints_season[9]]<-breakpoints_season[9]
  }
  if(percentil==0.02){
    values(raster_1_temp$yhat_c)[values(raster_1_temp$yhat_c)<=breakpoints[3]]<-breakpoints[3]
    values(raster_1_temp$yhat_c)[values(raster_1_temp$yhat_c)>=breakpoints[8]]<-breakpoints[8]

    values(raster_2_temp)[values(raster_2_temp)<=breakpoints_season[3]]<-breakpoints_season[3]
    values(raster_2_temp)[values(raster_2_temp)>=breakpoints_season[8]]<-breakpoints_season[8]
  }
  if(percentil==0.05){
    values(raster_1_temp$yhat_c)[values(raster_1_temp$yhat_c)<=breakpoints[4]]<-breakpoints[4]
    values(raster_1_temp$yhat_c)[values(raster_1_temp$yhat_c)>=breakpoints[7]]<-breakpoints[7]

    values(raster_2_temp)[values(raster_2_temp)<=breakpoints_season[4]]<-breakpoints_season[4]
    values(raster_2_temp)[values(raster_2_temp)>=breakpoints_season[7]]<-breakpoints_season[7]
  }
  if(percentil==0.10){
    values(raster_1_temp$yhat_c)[values(raster_1_temp$yhat_c)<=breakpoints[5]]<-breakpoints[5]
    values(raster_1_temp$yhat_c)[values(raster_1_temp$yhat_c)>=breakpoints[6]]<-breakpoints[6]

    values(raster_2_temp)[values(raster_2_temp)<=breakpoints_season[5]]<-breakpoints_season[5]
    values(raster_2_temp)[values(raster_2_temp)>=breakpoints_season[6]]<-breakpoints_season[6]
  }

  ### breaks para os gráficos na mesma escala ###
  cuts<-seq(min(minValue(raster_2_temp), minValue(raster_1_temp$yhat_c)),
    max(maxValue(raster_2_temp), maxValue(raster_1_temp$yhat_c)),length.out=100)
  cuts_plot<-round(seq(min(c(values(raster_1_temp$yhat_c), values(raster_2_temp))), na.rm=T),
    max(c(values(raster_1_temp$yhat_c), values(raster_2_temp))), na.rm=T),length.out=8),3) #set breaks

  #####
  ### Mapas na mesma escala ###
  #####
  png(paste0(dir_resultados, "plot_mesma_escala_3000_diario_", regiao, ".png"),
    width=4*680, height = 6*680, res=300)
    par(mfrow=c(3,2))
    par(mar=c(0,0,4,0), oma=c(0,0,0,0))
    plot(raster_2_temp$DJF, axes=F, breaks=cuts, box=F, col=colorRamps::matlab.like2(length(cuts)-1),
      legend=F,zlim=c(min(c(values(raster_2_temp), values(raster_1_temp)), na.rm=T),
        max(c(values(raster_2_temp), values(raster_1_temp))), na.rm=T)), main="DJF", cex.main=2)
    plot(contorno, add=T)

    plot(raster_2_temp$MAM, axes=F, breaks=cuts, box=F, col=colorRamps::matlab.like2(length(cuts)-1),
      legend=F,zlim=c(min(c(values(raster_2_temp), values(raster_1_temp)), na.rm=T),
        max(c(values(raster_2_temp), values(raster_1_temp))), na.rm=T)), main="MAM", cex.main=2)

```

```

plot(contorno, add=T)

plot(raster_2_temp$JJA, axes=F, breaks=cuts, box=F, col=colorRamps::matlab.like2(length(cuts)-1),
legend=F, zlim=c(min(c(values(raster_2_temp), values(raster_1_temp)), na.rm=T),
max(c(values(raster_2_temp), values(raster_1_temp)), na.rm=T)), main="JJA", cex.main=2)
plot(contorno, add=T)

plot(raster_2_temp$SON, axes=F, box=F, breaks=cuts, col=colorRamps::matlab.like2(length(cuts)-1),
legend=F, zlim=c(min(c(values(raster_2_temp), values(raster_1_temp)), na.rm=T),
max(c(values(raster_2_temp), values(raster_1_temp)), na.rm=T)), main="SON", cex.main=2)
plot(contorno, add=T)

plot(raster_1_temp$yhat_c, axes=F, box=F, breaks=cuts, col=colorRamps::matlab.like2(length(cuts)-1),
legend=F, zlim=c(min(c(values(raster_2_temp), values(raster_1_temp)), na.rm=T),
max(c(values(raster_2_temp), values(raster_1_temp)), na.rm=T)), main="ANUAL", cex.main=2)
plot(contorno, add=T)

plot(contorno, axes=F, lty=0)
par(new=T, xpd=T)
if(regiao=="RMSP"){
  legend_image <- rev(as.raster(matrix(colorRamps::matlab.like2(length(cuts)-1), ncol=1)))
  text(x=extent(rmspta)@xmin + (extent(rmspta)@xmax - extent(rmspta)@xmin)/2+30000,
y = seq(extent(rmspta)@ymin+8000, extent(rmspta)@ymax-8000, l=length(cuts_plot)),
labels = c(paste0("<=", round(cuts_plot[1], 3)), round(cuts_plot[2:(length(cuts_plot)-1)], 3),
paste0(">=", round(cuts_plot[length(cuts_plot)], 3))), xpd=NA, cex=2)
legend(legend=expression("MP"[2.5]*" ("*mu*"g/m^3*")"), x=extent(rmspta)@xmin +
(extent(rmspta)@xmax - extent(rmspta)@xmin)/2 - 16000, y=extent(rmspta)@ymax+15000,
bty="n", cex=2, xpd=NA)
rasterImage(legend_image, extent(rmspta)@xmin + (extent(rmspta)@xmax - extent(rmspta)@xmin)/2,
extent(rmspta)@ymin+6000, extent(rmspta)@xmin+(extent(rmspta)@xmax - extent(rmspta)@xmin)/2+8000,
extent(rmspta)@ymax-6000, xpd=NA)
}else{
  legend_image <- rev(as.raster(matrix(colorRamps::matlab.like2(length(cuts)-1), ncol=1)))
  text(x=extent(cidadeta)@xmin + (extent(cidadeta)@xmax - extent(cidadeta)@xmin)/2+20000,
y = seq(extent(cidadeta)@ymin+8000, extent(cidadeta)@ymax-8000, l=length(cuts_plot)),
labels = c(paste0("<=", round(cuts_plot[1], 3)), round(cuts_plot[2:(length(cuts_plot)-1)], 3),
paste0(">=", round(cuts_plot[length(cuts_plot)], 3))), xpd=NA, cex=2)
legend(legend=expression("MP"[2.5]*" ("*mu*"g/m^3*")"), x=extent(cidadeta)@xmin +
(extent(cidadeta)@xmax - extent(cidadeta)@xmin)/2 - 16000, y=extent(cidadeta)@ymax+6000, bty="n",
cex=2, xpd=NA)
rasterImage(legend_image, extent(cidadeta)@xmin + (extent(cidadeta)@xmax - extent(cidadeta)@xmin)/2,
extent(cidadeta)@ymin+6000, extent(cidadeta)@xmin+(extent(cidadeta)@xmax - extent(cidadeta)@xmin)/2+8000,
extent(cidadeta)@ymax-6000, xpd=NA)
}
dev.off()

#####
### Mapas em escalas diferentes ###
#####
png(paste0(dir_resultados, "plot_escala_diferente_3000_diario_", regiao, ".png"),
width=6*680, height = 6*680, res=300)
par(mfrow=c(3,2), mar=c(0,0,4,0), oma=c(3,2,2,9))
plot(raster_1_temp$yhat_c, col=pal(50), legend=F, axes=F, box=F, main="(A) Anual", cex.main=2.7)
plot(contorno, add=T)
plot(raster_1_temp$yhat_c, legend.only=TRUE, col=pal(50),
legend.width=1.8, legend.shrink=0.6,
axis.args=list(cex.axis=2),
legend.args=list(text=expression("MP"[2.5]*" ("*mu*"g/m^3*")"), side=3, font=2, cex=1.8, line=1.5),
axes=F, box=F)

plot(raster_1_temp$var_c, col=pal_erro(50), legend=F, axes=F, box=F, main="(B) Anual", cex.main=2.7)
plot(contorno, add=T)
plot(raster_1_temp$var_c, legend.only=TRUE, col=pal_erro(50),
legend.width=1.8, legend.shrink=0.6,
axis.args=list(cex.axis=2),
legend.args=list(text=expression("Desvio Padrão do \nErro de Predição ("*mu*"g/m^3*")"), side=3,
font=1, cex=1.6, line=1.5), axes=F, box=F)

plot(raster_2_temp$DJF, col=pal(50), legend=F, axes=F, box=F, main="(C) Verão", cex.main=2.7)
plot(contorno, add=T)
plot(raster_2_temp$DJF, legend.only=TRUE, col=pal(50),
legend.width=1.8, legend.shrink=0.6,
axis.args=list(cex.axis=2),
legend.args=list(text=expression("MP"[2.5]*" ("*mu*"g/m^3*")"), side=3, font=2, cex=1.8, line=1.5),
axes=F, box=F)

```

```
plot(raster_2_temp$MAM, col=pal(50), legend=F, axes=F, box=F, main="(D) Outono", cex.main=2.7)
plot(contorno, add=T)
plot(raster_2_temp$MAM, legend.only=TRUE, col=pal(50),
legend.width=1.8, legend.shrink=0.6,
axis.args=list(cex.axis=2),
legend.args=list(text=expression("MP"[2.5]*" ("*mu*"g/m"^3*")), side=3, font=2, cex=1.8, line=1.5),
axes=F, box=F)

plot(raster_2_temp$JJA, col=pal(50), legend=F, axes=F, box=F, main="(E) Inverno", cex.main=2.7)
plot(contorno, add=T)
plot(raster_2_temp$JJA, legend.only=TRUE, col=pal(50),
legend.width=1.8, legend.shrink=0.6,
axis.args=list(cex.axis=2),
legend.args=list(text=expression("MP"[2.5]*" ("*mu*"g/m"^3*")), side=3, font=2, cex=1.8, line=1.5),
axes=F, box=F)

plot(raster_2_temp$SON, col=pal(50), legend=F, axes=F, box=F, main="(F) Primavera", cex.main=2.7)
plot(contorno, add=T)
plot(raster_2_temp$SON, legend.only=TRUE, col=pal(50),
legend.width=1.8, legend.shrink=0.6,
axis.args=list(cex.axis=2),
legend.args=list(text=expression("MP"[2.5]*" ("*mu*"g/m"^3*")), side=3, font=2, cex=1.8, line=1.5),
axes=F, box=F)
dev.off()
```


E. *Script* para o ajuste do modelo que considera as duas fontes de correlação

A seguir apresentamos os comandos utilizados para ajustar o modelo E1A2 apresentado na seção 4.3.1. Para exemplificar, estamos considerando as concentrações diárias de $MP_{2.5}$. Também apresentamos os comandos utilizados na etapa de validação cruzada, na predição das concentrações, no cálculo dos desvios-padrão dos erros de predição e na confecção dos mapas de concentração.

```
#####  
### carrega as bibliotecas ###  
#####  
library(data.table)  
library(lme4)  
library(lmerTest)  
library(Matrix)  
library(gdata)  
library(raster)  
library(rgdal)  
library(lubridate)  
library(data.table)  
library(colorRamps)  
  
### definir localização dos dados e onde os resultados serão salvos  
dir_dados<-"~\\dados\\"  
dir_resultados<-"D:~\\resultados2\\"  
  
#####  
### leitura dos dados ###  
#####  
load(paste0(dir_dados, "dados_tese_completos_diario.RData"))  
  
#####  
### os dados estão no objeto "dados" ###  
#####  
head(dados) ### exibe as primeiras linhas de "dados"  
  
#####  
###seleciona apenas as variáveis de interesse ###  
#####  
variaveis<-c("DATA", "cell_number", "estacao_nome", "estacao_id", "nmp25", "temperatura", "umidade",  
"velocidade_vento", "precipitacao_acum", "road_density")  
  
temp_data<-dados[,..variaveis]  
rm(dados)  
  
#####  
### cria variáveis para o ajuste do modelo ###  
#####  
### cria a variável 'tempof'  
temp_data$tempof<-factor(as.numeric(temp_data$DATA)-15339)  
  
### cria a variável 'estacao_id' utilizada no ajuste do modelo  
temp_data$estacao_id<-factor(temp_data$estacao_id)  
  
### padroniza a variável densidade viária pelo valor máximo observado  
temp_data$density<-temp_data$road_density/max(temp_data$road_density, na.rm=T)  
  
### ordena os dados de acordo com o 'id' da estação  
temp_data<-temp_data[order(temp_data$estacao_id),]  
  
#####  
### separa os dados da etapa 1 ###
```

```
#####

### apenas as células com estação de monitoramento
dados_etapa_1<-temp_data[!is.na(temp_data$estacao_nome), ]
### ordena os dados pelo 'id' da estação e pela DATA
dados_etapa_1<-dados_etapa_1[order(dados_etapa_1$estacao_id, dados_etapa_1$DATA),]

#####
### ajuste do modelo ElA2 (Tabela 4.6) utilizando a função lmer do pacote lme4() ###
#####
fit<-lmerTest::lmer(nmp25~temperatura+precipitacao_acum+velocidade_vento+
(1|tempof)+(1|estacao_id),data=dados_etapa_1, na.action=na.omit, REML=F)
summary(fit)

#####
### Etapa de validação cruzada e cálculo das medidas de desempenho ###
#####
library(Matrix)
### função para a etapa de validação cruzada ###
funcao_CV<-function(s,dataset){
  dataset$id<-1:dim(dataset)[1]
  print(paste0("validação estação - ", s))
  dados_preditos_cv<-NULL
  cross_ajuste<-subset(dataset, cell_number!=s) ###dados ajuste
  cross_teste<-subset(dataset, cell_number==s) ###dados teste

  ### ajuste do modelo considerando a amostra sem os dados da estação s ###
  fit_ajuste<-lmer(fit, data=cross_ajuste, na.action=na.omit)

  ### parte aleatória do modelo ###
  ### independente da estrutura de variâncias-covariâncias escolhida,
  ### só conhecemos o efeito aleatório do tempo
  b<-ranef(fit_ajuste)$tempof[[1]]

  ### Matriz Zi ###
  ### tempos não observados na estação s ###
  no_obs<-as.numeric(colnames(Z)[1:length(b)][!(colnames(Z)[1:length(b)]%in%unique(cross_teste$tempof))])
  if(length(no_obs)==0){
    Zi<-Diagonal(n=length(b), x=1)
  }else{
    Zi<-Diagonal(n=length(b), x=1)
    Zi[, no_obs]<-0 ### zera os tempos não observados em uma determinada estação
  }

  aleathat<-Zi%*%b
  aleathat<-aleathat[-c(no_obs)] ### exclui os dias faltantes nessa estação

  ### Matriz Xi ###
  ### extrai as partes da formula do modelo
  f0<-gsub(' \\+ \\(1 \\| tempof\\)', "", gsub(' \\+ \\(1 \\| estacao_id\\)', "", formula(fit)))
  ### constroi a matriz do modelo, criando as dummies (intercepto)
  Xi<-model.matrix(formula(paste0(f0[2],"~", f0[3])), data=cross_teste)

  ### parte fixa do modelo ###
  betha<-summary(fit_ajuste)$coeff[,1]
  fixedhat<-Xi%*%t(t(betha))

  ### predição marginal ###
  yhat.m<-fixedhat

  ### predição condicional ###
  yhat.c<-fixedhat+aleathat

  ### armazena os dados preditos para a estação s ###
  dados_preditos_cv<-data.frame(cross_teste, yhat_c=as.matrix(yhat.c),
  yhat_m=as.matrix(yhat.m), aleat=as.matrix(aleathat))

  return(dados_preditos_cv)
}

#####
### Validação Cruzada ###
#####

data.set<-na.omit(dados_etapa_1)
```

```

rm(dados_etapa_1)

### número de tempos/dias
n_tempo<-ncol(table(drop.levels(as.factor(data.set$estacao_id)), drop.levels(as.factor(data.set$tempo))))

### Matrizes para o cálculo das predições e CV ###
#a lmer usa uma única matriz Z para os efeitos aleatórios;
M<-getME(fit,"Z")
Z<-M[,1:n_tempo] ### matriz 'Z' dos efeitos associados ao tempo

### Cross-Validadition - células com estação de monitoramento ###
mylist<-NULL
mylist<-lapply(unique(data.set$cell_number), dataset=data.set, funcao_CV)
dados_preditos<-NULL
dados_preditos<-do.call('rbind', mylist)

#####
### Cálculo das medidas de desempenho ###
#####

### Gráficos 'observado x predito' e medidas de desempenho por estação ###
tab_temp<-NULL
for(k in unique(dados_preditos$estacao_id)){
  dados_plot<-dados_preditos[dados_preditos$estacao_id==k,]

  print(paste0('desempenho estação - ', k))
  png(file=paste0(dir_resultados, k, "_observado_predito.png"), width=6*680, height=3*680, res=300)
  par(mfrow=c(1,1), mar=c(4,5,4,0))
  plot(dados_plot$DATA, dados_plot$ncmp25, type="l",
        ylab=expression("Média diária da concentração de MP"[2.5]*" ("*mu*"g/m^3*")"), xlab=" ",
        main=gsub("_", " ", paste0(unique(dados_plot$estacao_nome))),
        las=2, ylim=c(-5, max(dados_preditos$ncmp25, na.rm=T)), cex.main=1.3, cex.axis=1.3, cex.lab=1.3)
  lines(dados_plot$DATA, dados_plot$yhat_c, type="l", col='yellow', ylab=" ", cex=2)
  lines(dados_plot$DATA, dados_plot$yhat_m, type="l", col='red', ylab=" ", cex=2, lwd=2)
  legend("topleft", c("observado", "predição marginal", "predição condicional"), lty=1,
        col=c('black','red','yellow'), bty='n', cex=1.3)
  dev.off()

  EQMPc_temp<-sqrt(mean((dados_plot$ncmp25-dados_plot$yhat_c)^2, na.rm=T))
  R2c_temp<-summary(lm(dados_plot$ncmp25~dados_plot$yhat_c))$adj.r.squared

  tab_temp<-rbind(tab_temp, cbind(data.frame(unique(dados_plot$estacao_nome)), R2c_temp, EQMPc_temp))
}
### tabela com as medidas de desempenho de cada estação de monitoramento
tab_temp

### Medidas de desempenho global ###
R2_c<-summary(lm(dados_preditos$ncmp25~dados_preditos$yhat_c))$adj.r.squared
REQMP.c<-sqrt(mean((dados_preditos$ncmp25-dados_preditos$yhat_c)^2, na.rm=T))
round(summary(fit)$AICtab,4)[1]
rm(dados_preditos)

#####
### Etapa 2 com cálculo do desvio padrão do erro de predição (DPEP) ###
#####

### função para o cálculo das predições e do DPEP para cada célula ###
funcao_ET2_erro<-function(s,dataset){
  print(paste0("predicao célula - ", s))
  cross_teste<-subset(dataset, cell_number==s)

  ### Cálculo das predições e DPEP ###
  ### como para novas estações não conhecemos o efeito aleatório de estação (si)
  ### nas predições só vamos utilizar os efeitos aleatórios de 'tempo' (ut)
  ### mas o DPEP considera tanto a variância do 'tempo' quanto da 'estacao'

  ### independente do modelo só conhecemos o efeito aleatório do tempo
  b<-ranef(fit)$tempof[[1]]

  ### Matriz Zi
  ### tempos não observados na estação s ###
  no_obs<-colnames(Z)[1:length(b)][!(colnames(Z)[1:length(b)]%in%unique(cross_teste$tempof))]
  if(length(no_obs)==0){
    Zi<-Diagonal(n=length(b), x=1)
  }else{

```

```

    Zi<-Diagonal(n=length(b), x=1)
    Zi[, no_obs]<-0 ### zera os tempos não observados em uma determinada estação
  }

  ### Matriz Wi
  Wi<-t(t(rep(1, dim(Zi)[2])))

  ### parte aleatória do modelo
  ### independente do modelo só conhecemos o efeito aleatório do tempo
  aleathat<-Zi%*%b

  ### Matriz Xi ###
  ###extraí as partes da formula do modelo
  f0<-gsub(' \\+ \\(1 \\| tempof\\)', "", gsub(' \\+ \\(1 \\| estacao_id\\)', "", formula(fit)))
  ### constroi a matriz do modelo, criando as dummies (intercepto)
  Xi<-model.matrix(formula(paste0("~", f0[3])), data=cross_teste)

  ### parte fixa do modelo ###
  betha<-summary(fit)$coeff[,1]
  fixedhat<-Xi%*%t(t(betha))

  ### predição marginal ###
  yhat.m<-fixedhat

  ### predição condicional ###
  yhat.c<-fixedhat+aleathat

  ### variância do erro de predição marginal - as variâncias estão na diagonal ###
  var.m<-Xi%*%vcov(fit)%*%t(Xi) + 2*(Xi%*%cov_beta_b%*%t(Zi)) +
  Zi%*%G[1:length(b),1:length(b)]%*%t(Zi) + s2d*Wi%*%t(Wi) + Diagonal(n=dim(Xi)[1], x=s2)

  ### variância do erro de predição condicional - as variâncias estão na diagonal ###
  var.c<-Xi%*%vcov(fit)%*%t(Xi) + 2*(Xi%*%cov_beta_b%*%t(Zi)) + Zi%*%var_b%*%t(Zi) +
  s2d*Wi%*%t(Wi) + Diagonal(n=dim(Xi)[1], x=s2)

  dados_preditos_ET2<-NULL

  ### já estou guardando o DPEP
  dados_preditos_ET2<-cross_teste[, c("yhat_c", "var_c", "yhat_m", "var_m"):=list(as.matrix(yhat.c),
  as.matrix(sqrt(diag(var.c))),
  as.matrix(yhat.m),as.matrix(sqrt(diag(var.m))))]
  return(dados_preditos_ET2)
}

#####
### Matrizes e quantidades utilizadas nas predições e no DPEP ###
#####
### número de estações
n_id<-nrow(table(drop.levels(as.factor(data.set$estacao_id)), drop.levels(as.factor(data.set$tempo))))
rm(data.set)

### matriz do efeito associado à estação
rvcov_estacao_id<-VarCorr(fit)$estacao_id
rvcov_estacao_id<-attr(rvcov_estacao_id, "stddev")*attr(rvcov_estacao_id,
"correlation")*attr(rvcov_estacao_id, "stddev")

### matriz de Var-Cov do efeito associado ao tempo
rvcov_tempo<-VarCorr(fit)$tempof
rvcov_tempo<-attr(rvcov_tempo, "stddev")*attr(rvcov_tempo, "correlation")*attr(rvcov_tempo, "stddev")

### Matriz G ###
G_id<-kronecker(diag(n_id),rvcov_estacao_id)
G_tempo<-kronecker(diag(n_tempo),rvcov_tempo)
### ATENÇÃO: a ordem dos fatores depende do número de níveis;
### o com mais níveis aparece primeiro
G<-Matrix(bdiag(G_tempo,G_id), sparse=T)
rm(G_tempo)
rm(G_id)

### estimativa de sigma2, sigma2b e sigma2d ###
s2<-(attr(VarCorr(fit), "sc"))^2
s2d<-as.numeric(rvcov_estacao_id)
s2b<-as.numeric(rvcov_tempo)

### Matrizes para o cálculo dos DPEP ###

```

```

### Matriz 'W' dos efeitos associados às estações
W<-M[, (n_tempo+1):dim(M)[2]]
X<-getME(fit,"X") ### ordenado por estacao_id e DATA

#####
### Matrizes do Sistema de Equações de Henderson (A) ###
#####
m11<-(1/s2)*(t(X)%*%X)
m12<-(1/s2)*(t(X)%*%Z)
m13<-(1/s2)*(t(X)%*%W)
M1<-cbind(m11, m12, m13)
rm(m11)

m21<-(t(m12))
m22<-(1/s2)*(t(Z)%*%Z) + Matrix::diag(1/s2b, n_tempo)
m23<-(1/s2)*(t(Z)%*%W)
M2<-cbind(m21, m22, m23)
rm(m12)
rm(m21)
rm(m22)

m31<-(t(m13))
m32<-(t(m23))
m33<-(1/s2)*(t(W)%*%W) + Matrix::diag((1/s2d), n_id)
M3<-cbind(m31, m32, m33)
rm(m13)
rm(m31)
rm(m32)
rm(m33)

A<-rbind(M1, M2, M3)
beta<-summary(fit)$coeff[,1]
rm(M1)
rm(M2)
rm(M3)

### cálculo da variância - inversa da matriz do Sistema de Equações de Henderson (A)
invA<-solve(A)

### var(b)
var_b<-invA[(length(beta)+1):(length(beta)+length(ranef(fit)$tempof[[1]])),
(length(beta)+1):(length(beta)+length(ranef(fit)$tempof[[1]]))]

cov_beta_b<-invA[1:length(beta), (length(beta)+1):(length(beta)+length(ranef(fit)$tempof[[1]]))]
rm(A)
rm(invA)

### conjunto de dados completo
data.set<-temp_data[order(temp_data$cell_number),]
data.set<-data.set[which(!is.na(data.set$road_density)),]
data.set$id<-1:dim(data.set)[1]

### Cálculo das predições e dos DPEP para cada célula do grid###
### demorou cerca de 4h em um notebook com 16GB de RAM e processador Intel i7.
mylist<-NULL
mylist<-lapply(unique(data.set$cell_number), dataset=data.set, funcao_ET2_erro)
predicoes_erro_final<-NULL
print("rbind...")
predicoes_erro_final<-do.call('rbind', mylist)
rm(data.set)

### salva as predições e os respectivos DPEP em um arquivo ###
#save(predicoes_erro_final, file=paste0(dir_resultados, "predicoes_erro_final.RData"))

### Daqui em diante os comandos são idênticos aos apresentados no Apêndice E.

```


F. Predição das concentrações de MP_{2.5} e variância do erro de predição

Nas seções 1.7 e 3.3.3 apresentamos o modelo proposto pela *Metodologia de Harvard* e o modelo proposto na presente tese, respectivamente, utilizados para a predição das concentrações de MP_{2.5} na RMSP. Mas para facilitar o entendimento do cálculo da predição e da variância do erro de predição em cada modelo, considere os modelos simplificados descritos nas equações (F.1) e (F.2):

$$y_{it} = \beta_0 + \beta_1 x_{it} + b_t + \epsilon_{it}, i = 1, 2, \dots, n; t = 1, 2, \dots, t_i \quad (\text{F.1})$$

$$y_{it} = \beta_0 + \beta_1 x_{it} + b_t + s_i + \epsilon_{it}, i = 1, 2, \dots, n; t = 1, 2, \dots, t_i \quad (\text{F.2})$$

em que n é o número estações de monitoramento, t_i é o número de concentrações observadas na estação i , y_{it} é a concentração de MP_{2.5} observada na estação de monitoramento i no dia t , x_{it} é o valor observado da variável explicativa x no dia i no instante de tempo t (por exemplo, temperatura), β_0 e β_1 são os parâmetros associados aos efeitos fixos, b_t representa o efeito aleatório associado ao instante t , s_i representa o efeito aleatório associado à estação i e ϵ_{it} é o erro aleatório associado à observação realizada na estação i no dia t . Supõe-se $b_t \sim N(0, \sigma_b^2)$, $s_i \sim N(0, \sigma_s^2)$ e $\epsilon_{it} \sim N(0, \sigma^2)$, mutuamente independentes.

Considere ainda a situação em que temos 2 estações de monitoramento ($s = 2$) avaliadas ao longo de três dias ($t = 3$) cada. Nessa situação, podemos escrever (F.1) e (F.2) em notação matricial como apresentado nas equações (F.3) e (F.4).

$$Y = X\beta + Zb + \epsilon, \quad (\text{F.3})$$

em que

$$\underbrace{\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix}}_Y = \underbrace{\begin{bmatrix} 1 & x_{11} \\ 1 & x_{12} \\ 1 & x_{13} \\ 1 & x_{21} \\ 1 & x_{22} \\ 1 & x_{23} \end{bmatrix}}_X \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_Z \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}}_b + \underbrace{\begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \end{bmatrix}}_{\epsilon},$$

e supõe-se $b \sim N(0, G = \sigma_b^2 I_3)$ e $\epsilon \sim N(0, R = \sigma^2 I_6)$, mutuamente independentes.

$$Y = X\beta + Zb + Wd + \epsilon, \quad (\text{F.4})$$

$$\underbrace{\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix}}_Y = \underbrace{\begin{bmatrix} 1 & x_{11} \\ 1 & x_{12} \\ 1 & x_{13} \\ 1 & x_{21} \\ 1 & x_{22} \\ 1 & x_{23} \end{bmatrix}}_X \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_Z \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}}_b + \underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}}_W \underbrace{\begin{bmatrix} s_1 \\ s_2 \end{bmatrix}}_d + \underbrace{\begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \end{bmatrix}}_{\epsilon},$$

e supõe-se $b \sim N(0, G = \sigma_b^2 I_3)$, $d \sim N(0, D = \sigma_s^2 I_2)$ e $\epsilon \sim N(0, R = \sigma^2 I_6)$, mutuamente independentes.

O objetivo é prever as concentrações de $\text{MP}_{2.5}$ nos instantes $1, 2, \dots, t_k$ na célula k , que não possui estação de monitoramento, ou seja, queremos prever o vetor $y_k = (y_{k1}, y_{k2}, \dots, y_{kt_k})^\top$. Em (F.5) e (F.6) são apresentadas as equações dos modelos (F.1) e (F.2), respectivamente, em notação matricial, para a célula k .

$$y_k = X_k \beta + Z_k b + \epsilon_k, \quad (\text{F.5})$$

em que

$$\underbrace{\begin{bmatrix} y_{k1} \\ y_{k2} \\ y_{k3} \end{bmatrix}}_{y_k} = \underbrace{\begin{bmatrix} 1 & x_{k1} \\ 1 & x_{k2} \\ 1 & x_{k3} \end{bmatrix}}_{X_k} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{Z_k} \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}}_b + \underbrace{\begin{bmatrix} \epsilon_{k1} \\ \epsilon_{k2} \\ \epsilon_{k3} \end{bmatrix}}_{\epsilon_k}.$$

$$y_k = X_k \beta + Z_k b + W_k s_k + \epsilon_k, \quad (\text{F.6})$$

$$\underbrace{\begin{bmatrix} y_{k1} \\ y_{k2} \\ y_{k3} \end{bmatrix}}_{y_k} = \underbrace{\begin{bmatrix} 1 & x_{k1} \\ 1 & x_{k2} \\ 1 & x_{k3} \end{bmatrix}}_{X_k} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{Z_k} \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}}_b + \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}_{W_k} s_k + \underbrace{\begin{bmatrix} \epsilon_{k1} \\ \epsilon_{k2} \\ \epsilon_{k3} \end{bmatrix}}_{\epsilon_k}.$$

No contexto dos MLM existem dois tipos de predição: a condicional e marginal. A primeira leva em conta os efeitos aleatórios, e a segunda não. No contexto do problema que está sendo estudado, tanto no *Modelo de Harvard* quanto no modelo proposto na presente

tese, a predição condicional só pode ser condicionada nos efeitos aleatórios b_t , pois para as células sem estações de monitoramento, são os únicos que conseguimos calcular. Logo, para ambos os modelos as predições marginal (F.7) e condicional (F.8) são dadas por:

$$\hat{E}(y_k) = \hat{y}_k = X_k \hat{\beta} \quad (\text{F.7})$$

$$\hat{E}(y_k | \hat{b}) = \hat{y}_k | \hat{b} = X_k \hat{\beta} + Z_k \hat{b} \quad (\text{F.8})$$

Para o cálculo das predições pontuais apresentadas em F.7 e F.8 precisamos da estimativa de β ($\hat{\beta}$) e da predição dos efeitos aleatórios b ($\hat{b}$). Essas estimativas podem ser obtidas por meio do sistema de equações proposto por Henderson (1975), que no caso do modelo proposto na presente tese, fica definido como:

$$\underbrace{\begin{bmatrix} X^\top \hat{R}^{-1} X & X^\top \hat{R}^{-1} Z & X^\top \hat{R}^{-1} W \\ Z^\top \hat{R}^{-1} X & Z^\top \hat{R}^{-1} Z + \hat{G}^{-1} & Z^\top \hat{R}^{-1} W \\ W^\top \hat{R}^{-1} X & W^\top \hat{R}^{-1} Z & W^\top \hat{R}^{-1} W + \hat{D}^{-1} \end{bmatrix}}_A \begin{bmatrix} \hat{\beta} \\ \hat{b} \\ \hat{d} \end{bmatrix} = \begin{bmatrix} X^\top \hat{R}^{-1} Y \\ Z^\top \hat{R}^{-1} Y \\ W^\top \hat{R}^{-1} Y \end{bmatrix} \quad (\text{F.9})$$

Para o caso do *Modelo de Harvard*, basta desconsiderar em F.9 as componentes relacionadas à W . As equações que compõem o sistema de Henderson são as derivadas da função de distribuição conjunta dos efeitos aleatórios (b e d) e das observações Y ($f(Y, b, d)$), em relação à β , b e d . Mais detalhes podem ser encontrados em Singer et al. (2018), por exemplo.

Além das predições pontuais apresentadas em F.7 e F.8, precisamos apresentar uma medida de incerteza associada à essas predições¹. Essa etapa envolve o cálculo das variâncias de $\hat{\beta}$, $\hat{b}$, $\hat{d}$, e suas covariâncias, que também podem ser obtidas a partir do sistema de equações Henderson, como detalhado a seguir.

Considere a matriz A definida em F.9,

$$\begin{bmatrix} X^\top \hat{R}^{-1} X & X^\top \hat{R}^{-1} Z & X^\top \hat{R}^{-1} W \\ Z^\top \hat{R}^{-1} X & Z^\top \hat{R}^{-1} Z + \hat{G}^{-1} & Z^\top \hat{R}^{-1} W \\ W^\top \hat{R}^{-1} X & W^\top \hat{R}^{-1} Z & W^\top \hat{R}^{-1} W + \hat{D}^{-1} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{12}^\top & A_{22} & A_{23} \\ A_{13}^\top & A_{23}^\top & A_{33} \end{bmatrix} \quad (\text{F.10})$$

em que $\hat{R}$, $\hat{G}$ e $\hat{D}$ são as estimativas das matrizes de covariâncias R , G e D , respectivamente. Considere também a matriz inversa de A , definida como:

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{12}^\top & A_{22} & A_{23} \\ A_{13}^\top & A_{23}^\top & A_{33} \end{bmatrix}^{-1} = \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{12}^\top & C_{22} & C_{23} \\ C_{13}^\top & C_{23}^\top & C_{33} \end{bmatrix} \quad (\text{F.11})$$

¹ “The Statistician is the Wizard who makes “scientific” statements about invisible states and quantities. However, contrary to the real wishes (or witches), (s)he attaches uncertainties to (her) his statements.” Carlos Alberto de Bragança Pereira

Pode ser mostrado que:

- $Var(\hat{\beta}) = C_{11}$
- $Var(\hat{b}) = Cov(\hat{b}, b) = \hat{G} - C_{22}$
- $Var(\hat{b} - b) = C_{22}$
- $Cov(\hat{\beta}, \hat{b}) = 0$
- $Cov(\hat{\beta}, b) = -C_{12}$
- $Cov(\hat{\beta}, \hat{b} - b) = C_{12}$

e essas quantidades são utilizadas no cálculo da variância do erro de predição, detalhado a seguir.

F.1 Variância do erro de predição no Modelo de Harvard

F.1.1 Variância do erro de predição marginal

$$\begin{aligned}
Var(\hat{y}_k - y_k) &= Cov[(\hat{y}_k - y_k), (\hat{y}_k - y_k)^\top] = Cov[(X_k\hat{\beta} - (X_k\beta + Z_kb + \epsilon_k)), (X_k\hat{\beta} - (X_k\beta + Z_kb + \epsilon_k))^\top] \\
&= Cov[(X_k(\hat{\beta} - \beta) - (Z_kb + \epsilon_k)), (X_k(\hat{\beta} - \beta) - (Z_kb + \epsilon_k))^\top] \\
&= Cov[(X_k(\hat{\beta} - \beta)), (X_k(\hat{\beta} - \beta))^\top] \\
&\quad - 2Cov[X_k(\hat{\beta} - \beta), (Z_kb + \epsilon_k)^\top] + Cov[(Z_kb + \epsilon_k), (Z_kb + \epsilon_k)^\top] \\
&= Var[X_k(\hat{\beta} - \beta)] - 2Cov[X_k(\hat{\beta} - \beta), (Z_kb)^\top] - 2Cov[X_k(\hat{\beta} - \beta), (\epsilon_k)^\top] \\
&\quad + Cov[Z_kb, (Z_kb)^\top] + 2Cov[Z_kb, (\epsilon_k)^\top] + Cov(\epsilon_k, (\epsilon_k)^\top) \\
&= Var(X_k\hat{\beta}) - 2Cov[X_k\hat{\beta}, (Z_kb)^\top] - 0 + Var(Z_kb) + 0 + Var(\epsilon_k) \\
&= X_kVar(\hat{\beta})X_k^\top - 2X_kCov[\hat{\beta}, b]Z_k^\top + Z_kVar(b)Z_k^\top + Var(\epsilon_k) \\
&= X_kC_{11}X_k^\top - 2X_k[-C_{12}]Z_k^\top + Z_k\hat{G}Z_k^\top + \hat{R}
\end{aligned} \tag{F.12}$$

F.1.2 Variância do erro de predição condicional

$$\begin{aligned}
Var(\hat{y}_k - y_k | b = \hat{b}) &= Cov[(\hat{y}_k | \hat{b} - y_k), (\hat{y}_k | \hat{b} - y_k)^\top] \\
&= Cov[(X_k \hat{\beta} + Z_k \hat{b} - (X_k \beta + Z_k b + \epsilon_k)), (X_k \hat{\beta} + Z_k \hat{b} - (X_k \beta + Z_k b + \epsilon_k))^\top] \\
&= Cov[(X_k(\hat{\beta} - \beta) + Z_k(\hat{b} - b) - \epsilon_k), (X_k(\hat{\beta} - \beta) + Z_k(\hat{b} - b) - \epsilon_k)^\top] \\
&= Cov[(X_k(\hat{\beta} - \beta)), (X_k(\hat{\beta} - \beta))^\top] + 2Cov[X_k(\hat{\beta} - \beta), (Z_k(\hat{b} - b))^\top] \\
&\quad - 2Cov[X_k(\hat{\beta} - \beta), (\epsilon_k)^\top] + Cov[(Z_k(\hat{b} - b)), (Z_k(\hat{b} - b))^\top] - 2Cov[Z_k(\hat{b} - b), (\epsilon_k)^\top] \\
&\quad + Cov[(\epsilon_k), (\epsilon_k)^\top] \\
&= Var(X_k(\hat{\beta} - \beta)) + 2Cov[X_k(\hat{\beta} - \beta), (Z_k(\hat{b} - b))^\top] - 0 + Var[Z_k(\hat{b} - b)] - 0 + Var(\epsilon_k) \\
&= X_k Var(\hat{\beta}) X_k^\top + 2X_k Cov[\hat{\beta}, (\hat{b} - b)^\top] Z_k^\top + Z_k Var(\hat{b} - b) Z_k^\top + Var(\epsilon_k) \\
&= X_k C_{11} X_k^\top + 2X_k [C_{12}] Z_k^\top + Z_k [C_{22}] Z_k^\top + \hat{R}
\end{aligned} \tag{F.13}$$

F.2 Variância do erro de predição no modelo proposto

F.2.1 Variância do erro de predição marginal

$$\begin{aligned}
 Var(\hat{y}_k - y_k) &= Cov[(\hat{y}_k - y_k), (\hat{y}_k - y_k)^\top] \\
 &= Cov[(X_k \hat{\beta} - (X_k \beta + Z_k b + W_k s_k + \epsilon_k)), (X_k \hat{\beta} - (X_k \beta + Z_k b + W_k s_k + \epsilon_k))^\top] \\
 &= Cov[(X_k(\hat{\beta} - \beta) - (Z_k b + W_k s_k + \epsilon_k)), (X_k(\hat{\beta} - \beta) - (Z_k b + W_k s_k + \epsilon_k))^\top] \\
 &= Cov[X_k(\hat{\beta} - \beta), (X_k(\hat{\beta} - \beta))^\top] - 2Cov[X_k(\hat{\beta} - \beta), (Z_k b)^\top] - 2Cov[X_k(\hat{\beta} - \beta), (W_k s_k)^\top] \\
 &\quad - 2Cov[X_k(\hat{\beta} - \beta), (\epsilon_k)^\top] + Cov[Z_k b, (Z_k b)^\top] + 2Cov[Z_k b, (W_k s_k)^\top] + 2Cov[Z_k b, (\epsilon_k)^\top] \\
 &\quad + Cov[W_k s_k, (W_k s_k)^\top] + 2Cov[W_k s_k, (\epsilon_k)^\top] + Cov[\epsilon_k, (\epsilon_k)^\top] \\
 &= Var[X_k \hat{\beta}] - 2Cov[X_k \hat{\beta}, (Z_k b)^\top] - 2Cov[X_k \hat{\beta}, (W_k s_k)^\top] - 0 \\
 &\quad + Var[Z_k b] + 2Cov[Z_k b, (W_k s_k)^\top] + 0 + Var[W_k s_k] + 0 + Var[\epsilon_k] \\
 &= X_k Var(\hat{\beta}) X_k^\top - 2X_k Cov[\hat{\beta}, b] Z_k^\top - 2X_k Cov[\hat{\beta}, s_k] W_k^\top + Z_k Var(b) Z_k^\top \\
 &\quad + 2Z_k Cov(b, s_k) W_k^\top + W_k Var(s_k) W_k^\top + Var(\epsilon_k) \\
 &= X_k C_{11} X_k^\top - 2X_k [-C_{12}] Z_k^\top - 0^* + Z_k \hat{G} Z_k^\top + 0 + \hat{\sigma}_s^2 W_k W_k^\top + \hat{R}
 \end{aligned} \tag{F.14}$$

em que

$$\begin{aligned}
 {}^*Cov(\hat{\beta}, s_k) &= Cov[(X^\top V^{-1} X)^{-1} X^\top V^{-1} Y, s_k] = Cov(KY, s_k) = KCov(Y, s_k) \\
 &= KCov(X\beta + Zb + Wd + \epsilon, s_k) = KCov(Wd, s_k) \\
 &= KW Cov(d, s_k) = KW \begin{bmatrix} Cov(s_1, s_k) \\ Cov(s_2, s_k) \end{bmatrix} = KW \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},
 \end{aligned} \tag{F.15}$$

pois estamos assumindo $d \sim N(0, D)$, sendo $D = \sigma_s^2 I$. Ou seja, estamos assumindo que as estações são independentes.

F.2.2 Variância do erro de predição condicional

$$\begin{aligned}
Var(\hat{y}_k - y_k | b = \hat{b}) &= Cov[(\hat{y}_k - y_k | b = \hat{b}), (\hat{y}_k - y_k | b = \hat{b})^\top] \\
&= Cov[(X_k \hat{\beta} + Z_k \hat{b} - (X_k \beta + Z_k b + W_k s_k + \epsilon_k)), (X_k \hat{\beta} + Z_k \hat{b} - (X_k \beta + Z_k b + W_k s_k + \epsilon_k))^\top] \\
&= Cov[(X_k(\hat{\beta} - \beta) + Z_k(\hat{b} - b) - (W_k s_k + \epsilon_k)), (X_k(\hat{\beta} - \beta) + Z_k(\hat{b} - b) - (W_k s_k + \epsilon_k))^\top] \\
&= Cov[X_k(\hat{\beta} - \beta), (X_k(\hat{\beta} - \beta))^\top] + 2Cov[X_k(\hat{\beta} - \beta), (Z_k(\hat{b} - b))^\top] \\
&\quad - 2Cov[X_k(\hat{\beta} - \beta), (W_k s_k + \epsilon_k)^\top] + Cov[Z_k(\hat{b} - b), (Z_k(\hat{b} - b))^\top] \\
&\quad - 2Cov[Z_k(\hat{b} - b), (W_k s_k + \epsilon_k)^\top] + Cov[W_k s_k + \epsilon_k, (W_k s_k + \epsilon_k)^\top] \\
&= Var[X_k \hat{\beta}] + 2X_k Cov[\hat{\beta}, (\hat{b} - b)^\top] Z_k^\top - 2X_k Cov[\hat{\beta}, s_k^\top] W_k^\top - 0 + Var[Z_k(\hat{b} - b)] \\
&\quad - 2Cov[Z_k(\hat{b} - b), (W_k s_k)^\top] - 2Cov[Z_k(\hat{b} - b), (\epsilon_k)^\top] + Var(W_k s_k) + Var(\epsilon_k) \\
&= X_k C_{11} X_k^\top + 2X_k [C_{12}] Z_k^\top - 0^* + Z_k [C_{22}] Z_k^\top - 0 - 0 + \hat{\sigma}_s^2 W_k W_k^\top + \hat{R}
\end{aligned} \tag{F.16}$$

em que

$$^*Cov(\hat{b} - b, s_k) = Cov(\hat{b}, s_k) - Cov(b, s_k) = Cov(\hat{b}, s_k) = Cov(GZ^\top QY, s_k) = GZ^\top QCov(Y, s_k) = 0 \tag{F.17}$$

G. Concentrações diárias observadas e preditas pelo modelo E1A1

Figura G.1: Médias diárias da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A1 para a estação de monitoramento Pq. D. Pedro II em 2012-2017.

Figura G.2: Médias diárias da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A1 para a estação de monitoramento Santana em 2012-2017.

Figura G.3: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Moóca em 2012-2017.

Figura G.4: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Ibirapuera em 2012-2017.

Figura G.5: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Nossa Senhora do Ó em 2012-2017.

Figura G.6: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento São Caetano do Sul em 2012-2017.

Figura G.7: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Congonhas em 2012-2017.

Figura G.8: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Diadema em 2012-2017.

Figura G.9: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Santo Amaro em 2012-2017.

Figura G.10: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Osasco em 2012-2017.

Figura G.11: Médias diárias da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A1 para a estação de monitoramento Santo André Capuava em 2012-2017.

Figura G.12: Médias diárias da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A1 para a estação de monitoramento São Bernardo do Campo - Paulicéia em 2012-2017.

Figura G.13: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Taboão da Serra em 2012-2017.

Figura G.14: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Mauá em 2012-2017.

Figura G.15: Médias diárias da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A1 para a estação de monitoramento Pinheiros em 2012-2017.

Figura G.16: Médias diárias da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A1 para a estação de monitoramento Carapicuíba em 2012-2017.

Figura G.17: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Parelheiros em 2012-2017.

Figura G.18: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Cidade Universitária-USP em 2012-2017.

Figura G.19: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Santo André - Paço Municipal em 2012-2017.

Figura G.20: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Itaim Paulista em 2012-2017.

Figura G.21: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Interlagos em 2012-2017.

Figura G.22: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Guarulhos - Paço Municipal em 2012-2017.

Figura G.23: Médias diárias da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A1 para a estação de monitoramento Marginal Tietê - Ponte dos Remédios em 2012-2017.

Figura G.24: Médias diárias da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A1 para a estação de monitoramento Capão Redondo em 2012-2017.

Figura G.25: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento São Bernardo - Centro em 2012-2017.

Figura G.26: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Guarulhos - Pimentas em 2012-2017.

H. Concentrações diárias observadas e preditas pelo modelo E1A2

Figura H.1: Médias diárias da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A2 para a estação de monitoramento Pq. D. Pedro II em 2012-2017.

Figura H.2: Médias diárias da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A2 para a estação de monitoramento Santana em 2012-2017.

Figura H.3: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Moóca em 2012-2017.

Figura H.4: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Ibirapuera em 2012-2017.

Figura H.5: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Nossa Senhora do Ó em 2012-2017.

Figura H.6: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento São Caetano do Sul em 2012-2017.

Figura H.7: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Congonhas em 2012-2017.

Figura H.8: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Diadema em 2012-2017.

Figura H.9: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Santo Amaro em 2012-2017.

Figura H.10: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Osasco em 2012-2017.

Figura H.11: Médias diárias da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A2 para a estação de monitoramento Santo André Capuava em 2012-2017.

Figura H.12: Médias diárias da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A2 para a estação de monitoramento São Bernardo do Campo - Paulicéia em 2012-2017.

Figura H.13: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Taboão da Serra em 2012-2017.

Figura H.14: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Mauá em 2012-2017.

Figura H.15: Médias diárias da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A2 para a estação de monitoramento Pinheiros em 2012-2017.

Figura H.16: Médias diárias da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A2 para a estação de monitoramento Carapicuíba em 2012-2017.

Figura H.17: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Parelheiros em 2012-2017.

Figura H.18: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Cidade Universitária-USP em 2012-2017.

Figura H.19: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Santo André - Paço Municipal em 2012-2017.

Figura H.20: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Itaim Paulista em 2012-2017.

Figura H.21: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Interlagos em 2012-2017.

Figura H.22: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Guarulhos - Paço Municipal em 2012-2017.

Figura H.23: Médias diárias da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A2 para a estação de monitoramento Marginal Tietê - Ponte dos Remédios em 2012-2017.

Figura H.24: Médias diárias da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A2 para a estação de monitoramento Capão Redondo em 2012-2017.

Figura H.25: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento São Bernardo - Centro em 2012-2017.

Figura H.26: Médias diárias da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Guarulhos - Pimentas em 2012-2017.

I. Concentrações diurnas observadas e preditas pelo modelo E1A1

Figura I.1: Médias diurnas da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A1 para a estação de monitoramento Pq. D. Pedro II em 2012-2017.

Figura I.2: Médias diurnas da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A1 para a estação de monitoramento Santana em 2012-2017.

Figura I.3: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Moóca em 2012-2017.

Figura I.4: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Ibirapuera em 2012-2017.

Figura I.5: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Nossa Senhora do Ó em 2012-2017.

Figura I.6: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento São Caetano do Sul em 2012-2017.

Figura I.7: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Congonhas em 2012-2017.

Figura I.8: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Diadema em 2012-2017.

Figura I.9: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Santo Amaro em 2012-2017.

Figura I.10: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Osasco em 2012-2017.

Figura I.11: Médias diurnas da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A1 para a estação de monitoramento Santo André Capuava em 2012-2017.

Figura I.12: Médias diurnas da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A1 para a estação de monitoramento São Bernardo do Campo - Paulicéia em 2012-2017.

Figura I.13: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Taboão da Serra em 2012-2017.

Figura I.14: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Mauá em 2012-2017.

Figura I.15: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Pinheiros em 2012-2017.

Figura I.16: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Carapicuíba em 2012-2017.

Figura I.17: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Parelheiros em 2012-2017.

Figura I.18: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Cidade Universitária-USP em 2012-2017.

Figura I.19: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Santo André - Paço Municipal em 2012-2017.

Figura I.20: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Itaim Paulista em 2012-2017.

Figura I.21: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Interlagos em 2012-2017.

Figura I.22: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Guarulhos - Paço Municipal em 2012-2017.

Figura I.23: Médias diurnas da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A1 para a estação de monitoramento Marginal Tietê - Ponte dos Remédios em 2012-2017.

Figura I.24: Médias diurnas da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A1 para a estação de monitoramento Capão Redondo em 2012-2017.

Figura I.25: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento São Bernardo - Centro em 2012-2017.

Figura I.26: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A1 para a estação de monitoramento Guarulhos - Pimentas em 2012-2017.

J. Concentrações diurnas observadas e preditas pelo modelo E1A2

Figura J.1: Médias diurnas da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A2 para a estação de monitoramento Pq. D. Pedro II em 2012-2017.

Figura J.2: Médias diurnas da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A2 para a estação de monitoramento Santana em 2012-2017.

Figura J.3: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Moóca em 2012-2017.

Figura J.4: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Ibirapuera em 2012-2017.

Figura J.5: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Nossa Senhora do Ó em 2012-2017.

Figura J.6: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento São Caetano do Sul em 2012-2017.

Figura J.7: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Congonhas em 2012-2017.

Figura J.8: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Diadema em 2012-2017.

Figura J.9: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Santo Amaro em 2012-2017.

Figura J.10: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Osasco em 2012-2017.

Figura J.11: Médias diurnas da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A2 para a estação de monitoramento Santo André Capuava em 2012-2017.

Figura J.12: Médias diurnas da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A2 para a estação de monitoramento São Bernardo do Campo - Paulicéia em 2012-2017.

Figura J.13: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Taboão da Serra em 2012-2017.

Figura J.14: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Mauá em 2012-2017.

Figura J.15: Médias diurnas da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A2 para a estação de monitoramento Pinheiros em 2012-2017.

Figura J.16: Médias diurnas da concentração de $MP_{2.5}$ observadas e preditas pelo modelo E1A2 para a estação de monitoramento Carapicuíba em 2012-2017.

Figura J.17: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Parelheiros em 2012-2017.

Figura J.18: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Cidade Universitária-USP em 2012-2017.

Figura J.19: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Santo André - Paço Municipal em 2012-2017.

Figura J.20: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Itaim Paulista em 2012-2017.

Figura J.21: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Interlagos em 2012-2017.

Figura J.22: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Guarulhos - Paço Municipal em 2012-2017.

Figura J.23: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Marginal Tietê - Ponte dos Remédios em 2012-2017.

Figura J.24: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Capão Redondo em 2012-2017.

Figura J.25: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento São Bernardo - Centro em 2012-2017.

Figura J.26: Médias diurnas da concentração de $MP_{2.5}$ observadas e previstas pelo modelo E1A2 para a estação de monitoramento Guarulhos - Pimentas em 2012-2017.

K. Mapas das concentrações de $MP_{2.5}$ para a RMSP representados na mesma escala

Figura K.1: Média da concentração diurna de $MP_{2.5}$ predita pelo modelo E1A1 para a RMSP em 2012-2017 e de acordo com as estações do ano. DJF: Dezembro, Janeiro, Fevereiro; MAM: Março, Abril, Maio; JJA: Junho, Julho, Agosto, SON: Setembro, Outubro, Novembro.

Figura K.2: Média da concentração diária de $MP_{2.5}$ predita pelo modelo E1A1 para a RMSP em 2012-2017 e de acordo com as estações do ano. DJF: Dezembro, Janeiro, Fevereiro; MAM: Março, Abril, Maio; JJA: Junho, Julho, Agosto, SON: Setembro, Outubro, Novembro.

Figura K.3: Média da concentração diurna de $MP_{2.5}$ predita pelo modelo E1A2 para a RMSP em 2012-2017 e de acordo com as estações do ano. DJF: Dezembro, Janeiro, Fevereiro; MAM: Março, Abril, Maio; JJA: Junho, Julho, Agosto, SON: Setembro, Outubro, Novembro.

Figura K.4: Média da concentração diária de $MP_{2.5}$ predita pelo modelo E1A2 para a RMSP em 2012-2017 e de acordo com as estações do ano. DJF: Dezembro, Janeiro, Fevereiro; MAM: Março, Abril, Maio; JJA: Junho, Julho, Agosto, SON: Setembro, Outubro, Novembro.

L. Mapas das concentrações de $MP_{2.5}$ para a Cidade de São Paulo

Figura L.1: Médias anual (A) e sazonais (C-F) das concentrações diurnas de $MP_{2.5}$ previstas pelo modelo E1A1 para a Cidade de São Paulo em 2012-2017. (B) Média dos desvios-padrão dos erros de predição. Para uma melhor visualização os gráficos estão em escalas diferentes.

Figura L.2: Médias anual (A) e sazonais (C-F) das concentrações diurnas de $MP_{2.5}$ preditas pelo modelo E1A2 para a Cidade de São Paulo em 2012-2017. (B) Média dos desvios-padrão dos erros de predição. Para uma melhor visualização os gráficos estão em escalas diferentes.

Figura L.3: Médias anual (A) e sazonais (C-F) das concentrações diárias de $MP_{2.5}$ preditas pelo modelo E1A1 para a Cidade de São Paulo em 2012-2017. (B) Média dos desvios-padrão dos erros de predição. Para uma melhor visualização os gráficos estão em escalas diferentes.

Figura L.4: Médias anual (A) e sazonais (C-F) das concentrações diárias de $MP_{2.5}$ preditas pelo modelo E1A2 para a Cidade de São Paulo em 2012-2017. (B) Média dos desvios-padrão dos erros de predição. Para uma melhor visualização os gráficos estão em escalas diferentes.

M. Mapas das concentrações de $MP_{2.5}$ para a Cidade de São Paulo representados na mesma escala

Figura M.1: Média da concentração diurna de $MP_{2.5}$ predita pelo modelo E1A1 para a Cidade de São Paulo em 2012-2017 e de acordo com as estações do ano. DJF: Dezembro, Janeiro, Fevereiro; MAM: Março, Abril, Maio; JJA: Junho, Julho, Agosto, SON: Setembro, Outubro, Novembro.

Figura M.2: Média da concentração diária de $MP_{2.5}$ predita pelo modelo E1A1 para a Cidade de São Paulo em 2012-2017 e de acordo com as estações do ano. DJF: Dezembro, Janeiro, Fevereiro; MAM: Março, Abril, Maio; JJA: Junho, Julho, Agosto, SON: Setembro, Outubro, Novembro.

Figura M.3: Média da concentração diurna de $MP_{2.5}$ predita pelo modelo E1A2 para a Cidade de São Paulo em 2012-2017 e de acordo com as estações do ano. DJF: Dezembro, Janeiro, Fevereiro; MAM: Março, Abril, Maio; JJA: Junho, Julho, Agosto, SON: Setembro, Outubro, Novembro.

Figura M.4: Média da concentração diária de $MP_{2.5}$ predita pelo modelo E1A2 para a Cidade de São Paulo em 2012-2017 e de acordo com as estações do ano. DJF: Dezembro, Janeiro, Fevereiro; MAM: Março, Abril, Maio; JJA: Junho, Julho, Agosto, SON: Setembro, Outubro, Novembro.

ANEXOS

1. Artigo publicado na Revista
Atmospheric Environment

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: <http://www.elsevier.com/locate/atmosenv>

Exploring the relationship between high-resolution aerosol optical depth values and ground-level particulate matter concentrations in the Metropolitan Area of São Paulo

Aline Santos Damascena^{a,*}, Márcia Akemi Yamasoe^b, Vitor Souza Martins^c, Jorge Rosas^b, Noelia Rojas Benavente^b, Maciel Piñero Sánchez^b, Nelson Ithiro Tanaka^d, Paulo Hilário Nascimento Saldiva^a

^a Departamento de Patologia, Faculdade de Medicina, Universidade de São Paulo, Avenida Doutor Arnaldo, 455, São Paulo, 01246-903, Brazil

^b Departamento de Ciências Atmosféricas, Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, Rua do Matão, 1226, Cidade Universitária, 05508-090, Brazil

^c Agricultural and Biosystems Engineering, Iowa State University, 1340 Elings Hall, 605 Bissell Road, Ames, IA, 50010, USA

^d Departamento de Estatística, Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão, 1010, Cidade Universitária, 05508-090, Brazil

HIGHLIGHTS

- MAIAC AOD retrievals correlated weakly with PM concentrations in the MASP.
- We evaluated MAIAC AOD retrievals over the MASP.
- The MAIAC aerosol model is not compatible with the aerosol properties in the MASP.
- There is a wide variety of aerosol types in the MASP.

ARTICLE INFO

Keywords:

Moderate Resolution Imaging Spectroradiometer
Multi-Angle Implementation of Atmospheric Correction
Particulate matter
Air pollution
Metropolitan Area of São Paulo
Brazil

ABSTRACT

The spatiotemporal pattern of particulate matter (PM) concentrations is an important factor in predicting health issues in inhabitants of urban areas. The integration of satellite-derived aerosol optical depth (AOD) data with ground-level PM concentration data, obtained from monitoring networks, has contributed to better characterization of the spatiotemporal variability of aerosols worldwide. However, before using satellite AOD data as a proxy for PM in epidemiological and air quality studies in specific regions, the applicability of that strategy must be evaluated. In this study, we evaluate the use of the high-resolution AOD, derived from Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm, as a predictor of surface PM concentrations in the Metropolitan Area of São Paulo (MASP). We found relatively weak or negative correlations between PM concentrations and MAIAC AOD, even after vertical correction by planetary boundary layer height and the hygroscopic growth factor. The weak correlations reported in this study are mainly due to the mismatch between the current MAIAC aerosol model and the properties of local aerosols in the MASP. Our results suggest that sources of aerosol particles in the MASP are quite diverse and that there is therefore no single optical model suitable for use with satellite-derived AOD.

1. Introduction

Particulate matter (PM) is the solid component of atmospheric aerosol, which is defined as a suspension of solid or liquid particles in a

gas (Seinfeld and Pandis, 2016). These particles play a central role in regulating climate and air quality, by directly and indirectly affecting the radiative balance on Earth. Depending on their size and composition, these particles can scatter or absorb solar radiation, changing the surface

* Corresponding author. Departamento de Patologia, Faculdade de Medicina, Universidade de São Paulo, Avenida Doutor Arnaldo, 455, Cerqueira César, SP 01246-903, Brazil.

E-mail addresses: asantosdamascena@gmail.com, aline.damascena@usp.br (A.S. Damascena).

<https://doi.org/10.1016/j.atmosenv.2020.117949>

Received 7 May 2020; Received in revised form 14 August 2020; Accepted 18 September 2020

Available online 20 September 2020

1352-2310/© 2020 Elsevier Ltd. All rights reserved.

temperature (direct effect) or altering the cloud lifetime and albedo (indirect effect) (IPCC, 2013). In addition, the impact of PM on human health is well known, the degree of harm being inversely related to the particle size. A literature review correlating PM₁₀ (particles with aerodynamic diameter of less than 10 µm) and PM_{2.5} (particles with aerodynamic diameter of less than 2.5 µm) with various morbidity and mortality outcomes is presented in a World Health Organization (WHO) report (World Health Organization, 2006). In this article, PM refers only to PM_{2.5} and PM₁₀ and no other size of particulate matter.

The most harmful air pollutants, such as PM, are typically measured in air quality monitoring networks, which are sparse in low- and middle-income countries (Fajersztajn et al., 2014; World Health Organization, 2016). In Brazil, there are 27 “federative units” (26 states and the Federal District of Brasília), of which only eight have air pollutant monitoring networks: Bahia, Espírito Santo, Federal District of Brasília, Minas Gerais, Paraná, Rio de Janeiro, Rio Grande do Sul and São Paulo. Of those eight, only the states of São Paulo and Rio de Janeiro have monitoring networks somewhat comparable to those employed in the United States and Europe (Alves et al., 2014; IEMA, 2014). In view of that, the integration of high-resolution satellite data with ground-level monitoring network data can contribute to better characterization of the spatiotemporal variability of aerosols at local and regional scales (Chudnosvsky et al., 2014a).

The satellite-retrieved parameter related to PM is the Aerosol Optical Depth (AOD), which is a quantitative measure of the attenuation of the electromagnetic radiation by the presence of aerosols in the atmospheric column, at a specific wavelength (Wallace and Hobbs, 2006). The AOD is one of the atmospheric products of the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor aboard the Terra and Aqua satellites. However, the coarse spatial resolution of the AOD (10 × 10 km² and 3 × 3 km²) has prohibited its applicability in local scale studies. Recently, a higher spatial resolution (1 × 1 km²) MODIS AOD was derived from the Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm (Lyapustin et al., 2018), which can have beneficial applications in epidemiological and air quality studies, mainly in urban regions.

In the past 10–15 years, satellite-retrieved AOD data have been successfully applied in several approaches, in order to retrospectively predict the concentrations of PM, which can be used by health studies. Those approaches can be classified as statistical or geophysical (Jin et al., 2019). Chu et al. (2016) presented a review of the studies that used different methodologies for the retrospective prediction of PM_{2.5} using satellite-retrieved AOD. The authors included 116 studies and found that four types of model were the most frequently used in the predictions: Multiple Linear, Linear Mixed Model, Geographically Weighted Regression and Chemical Transport Model. The last is a geophysical approach while the others are statistical. Examples of the application of these approaches to retrospectively predict PM₁₀ can be found in the works of, amongst others, Stafoggia et al. (2017) and Kumar et al. (2016). Recently, in order to retrospectively predict the PM concentrations, there have also been approaches that combine the geophysical and statistical approaches (e.g., Knibbs et al., 2018; Beloconi et al., 2018; Meng et al., 2019; van Donkelaar et al., 2019).

The populations most affected by air pollution are those living in urban areas, in which pollutant levels often exceed the standards established by the WHO (World Health Organization, 2016). In 2019, about 56% of the global population resided in urban areas (World Urbanization Prospects, 2019a), while in the South America that percentage was higher than 85% (World Urbanization Prospects, 2019b). The Metropolitan Area of São Paulo (MASP) is the most populous region in Latin America, and vehicle emissions constitute the main source of air pollution in the area (CETESB, 2017). Although the effects of air pollution levels on human health in the MASP have been described in many epidemiological studies (Braga et al., 2001; Ribeiro and Cardoso, 2003; Saldiva et al., 1994, 1995; Ribeiro et al., 2019; Takano et al., 2019), few studies have attempted to retrospectively predict surface PM concentrations (Habermann and Gouveia, 2012), and only one study has

investigated the applicability of satellite-retrieved AOD (Natali, 2008), in this area.

Before using satellite AOD data as a proxy for PM in epidemiological studies and air quality studies in specific regions, the applicability of the strategy must be evaluated. Therefore, the aim of this study was to evaluate the use of high-resolution MAIAC AOD data to predict surface PM concentrations in the MASP in the 2012–2017 period. To that end, we integrated ground-level measurements with satellite data, describing the data processing procedures. In addition, we presented an evaluation of the MAIAC-AODs data performance over the MASP.

2. Methods

This study was conducted in the MASP, which comprises the city of São Paulo and 39 surrounding cities (Fig. 1). It is an urban region with a total area of 7946 km² and over 21 million inhabitants (<https://emplasa.sp.gov.br/RMSP>).

The climate of the MASP is characterized by two seasons (Andrade et al., 2012a): a wet season (between October and March) and a dry season (between April and September). During the dry season, when the planetary boundary layer (PBL) is low and atmospheric conditions do not favor air pollutant dispersion, episodes of high PM concentrations are common in the MASP (de Almeida Castanho and Artaxo, 2001; Albuquerque et al., 2012; de Miranda et al., 2012; Carvalho et al., 2015; CETESB, 2017).

2.1. Particulate matter surface data at MASP

Hourly PM mass concentrations data, in the 2012–2017 period, were obtained from automatic air quality monitoring stations operated by the São Paulo State *Companhia de Tecnologia de Saneamento Ambiental* (CETESB, Environmental Protection Agency). The CETESB measures PM concentrations using an automated beta attenuation monitor (CETESB, 2017; Xavier, 2011; Thermo Fisher Scientific Inc., 2010), with a minimum detectable concentration limit of 4 µg/m³ and a 1-h time resolution.

In the MASP, CETESB stations are impacted mainly by vehicular emissions. The difference is in the fleet type, although some of the stations are located near industrial areas (Mauá and Santo André-Capuava) or inside parks (Ibirapuera, Cidade Universitária - USP and Pico do Jaraguá). The stations at Congonhas, Marginal Tietê-Ponte dos Remédios, Osasco, Pinheiros e São Caetano do Sul are impacted by light-duty and heavy-duty vehicles. The others are affected mainly by light-duty vehicles and buses (Carvalho et al., 2015).

Considering the recommendations of the “Code of Federal Regulations - Title 40 - Protection of Environment/Part 58 - Ambient Air Quality Surveillance/Appendix E to Part 58 - Probe and Monitoring Path Siting Criteria for Ambient Air Quality Monitoring” developed by the United States Environmental Protection Agency (USEPA), CETESB classifies the monitoring stations according to their spatial representativeness. The classification takes into account the pollutant, the characteristics of the monitored region (e.g., topographic and meteorological characteristics), as well as the impact of emissions from fixed and mobile sources (e.g., proximity and intensity) in the stations (CETESB, 2016).

Regarding the PM concentrations, since in the MASP the main source of these concentrations is vehicular emission, the distance from the major traffic lanes is one of the characteristics considered in the spatial representativeness classification. The stations are classified on the micro, medium, suburban and urban scales. The former two are highly impacted by the local emissions while the latter are representative of areas of dimensions varying between 501 m and 50 km (CETESB, 2016).

In the present study, we considered the stations classified as suburban and urban, since their spatial representativeness is close to or higher than that of the satellite data. The number of CETESB stations measuring PM concentrations, which met the criteria for spatial representativeness in the present study, varied during the study period (Table S1 in the

Fig. 1. Topography map of the study area showing the city of São Paulo (SP city), as well as the other cities in the MASP, and the locations of the CETESB monitoring stations used in this study in 2017, the AERONET site, and the Campo de Marte Airport. Red and orange dots represent stations measuring PM_{2.5} and PM₁₀ only, respectively. Green dots represent stations measuring both pollutants. Inset shows the state of São Paulo (in red) within a map of Brazil. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Supplementary Information (SI)). Fig. 1 shows the location of the ground-level monitoring stations used in the present study for the year 2017.

For each CETESB station, we considered three distinct time intervals for calculating the mean PM concentrations: the hour of the satellite overpasses; diurnal (average of measurements obtained over 6–18 h periods); and daily (average of 24-h measurements). Diurnal and daily PM mean concentrations were included because they are used in epidemiological studies.

2.2. MODIS–MAIAC AOD

In this study, we used the MODIS–MAIAC AOD (MCD19A2 product), with $1 \times 1 \text{ km}^2$ of spatial resolution, at a wavelength of 550 nm, retrieved from the Terra and Aqua satellites for the 2012–2017 period. The data are available on Earthdata website (<https://earthdata.nasa.gov>), and we used only the high-quality retrievals (those with a Quality Assurance bit 0000, which represents “Best quality”; https://modisland.gsfc.nasa.gov/pdf/MCD19_UserGuide_final_Feb-6-2018.pdf). The data were processed by using the HDF-EOS to GeoTIFF Conversion Tool (HEG, <https://newsroom.gsfc.nasa.gov/sdptoolkit/HEG/HEGHome.html>). During the study period, the Terra satellite passed over the MASP from 9:00 to 11:30 (local time), whereas the Aqua satellite passed over the area from 12:45 to 15:05 (local time).

A detailed description of the MAIAC algorithm can be found in the studies conducted by Lyapustin et al. (2011a, b; 2012; 2018). In brief, it is a generic algorithm based on image-processing and time series analysis. Validation against AEROSOL ROBOTIC NETWORK (AERONET, <https://aeronet.gsfc.nasa.gov>; Holben et al., 1998) AOD data has shown that the MAIAC algorithm is capable of retrieving aerosol data over dark (vegetated) surfaces with accuracy comparable to that of the MODIS level 2 aerosol product from the Terra satellite ($10 \times 10 \text{ km}^2$ resolution) and that its accuracy increases over brighter surfaces, such as deserts and urban areas in the United States (Lyapustin et al., 2011b). The validation study conducted by Zhang et al. (2019) over China, showed that the MAIAC performed better in urban/cropland regions than in other regions (arid, semiarid, general vegetation, and water).

The MAIAC AOD data (Terra and Aqua retrievals) have been validated across South America. The validation procedure consisted of comparing the average MAIAC AOD retrievals collocated spatially and temporally with the average AOD from AERONET sites in the 2000–2016 period. The Pearson’s correlation coefficient (R) for the relationships that the AOD data from the São Paulo AERONET station, located atop the tallest building of the Physics Institute on the main

campus of the University of São Paulo (latitude: -23.561° , longitude: -46.735°), had with the Terra and Aqua retrievals of MAIAC AOD data was $R = 0.90$ and $R = 0.82$, respectively. Details about the validation study of MAIAC data for South America can be found in Martins et al. (2017).

2.3. Corrections by PBL height and by the hygroscopic growth factor

The relationship between AOD and PM concentrations (AOD–PM relationship) is not straightforward and is affected by a variety of factors, including the PBL height, the relative humidity (RH), and the vertical distribution of aerosols (Chu et al., 2003; Li et al., 2015). Because AOD is a value associated with the atmospheric column, it is affected by the PBL height, which has an impact on the vertical profile of aerosols. In general, aerosols are well mixed under the PBL (Gupta et al., 2006; Koelemeijer et al., 2006; Guinot et al., 2006; Zhang et al., 2009a). However, when there is aerosol transport at different heights, the AOD–PM relationship is altered, because ground-level monitoring stations do not capture the PM concentrations at higher levels. In addition, disconnected aerosol layers in the atmospheric column can differ in terms of their physical and chemical composition (Slater and Dibb, 2004). Therefore, the PBL height and the vertical distribution of aerosols are key variables for determining the AOD–PM relationship.

Assuming that aerosols are well mixed under the PBL, previous studies have corrected satellite-derived AOD using the vertical correction depicted in equation (1) (Koelemeijer et al., 2006; Schaap et al., 2009; Boyouk et al., 2010; Chu et al., 2013; Tsai et al., 2011; Chew et al., 2016; Kong et al., 2016; Zheng et al., 2017; Gong et al., 2017):

$$ASH = \frac{AOD}{PBLH} \quad (1)$$

where ASH is the aerosol scale height and PBLH is the PBL height. Given that we are not integrating the aerosol extinction in the total atmospheric column, the term “AOD” is incorrect and we will use the term “ASH”. That correction eliminates aerosols at higher levels, taking into account only those under the PBL, thereby typically improving the AOD–PM relationship (Koelemeijer et al., 2006).

We calculated morning and afternoon estimates of the PBL height using balloon sounding data from Campo de Marte Airport (latitude: -23.515° , longitude: -46.647°). Atmospheric balloon sounding data for stations around the world are available in <http://weather.uwyo.edu/upperair/sounding.html>, balloons being launched daily at 12:00 UTC and 00:00 UTC. Following the methodology employed by Sánchez et al. (2020), who studied the dynamic and thermodynamic properties of the

PBL in the MASP, we estimated morning and afternoon PBL height by the parcel and air temperature gradient methods, respectively.

We obtained morning PBL height estimates by virtual potential temperature profiles from 12:00 UTC balloon sounding data using the parcel method, in which the estimated PBL height is that at which the virtual potential temperature is the same as the surface virtual potential temperature. Due to the fact that balloon soundings are not carried out during the afternoon, we estimated the residual-mixing layer (RML) height by determining the air temperature profiles at 00:00 UTC and, as suggested by Sánchez et al. (2020), used the RML height as an approximation of the maximum daytime PBL height. The RML height was obtained, by using the air temperature gradient method, as the height at which the air temperature gradient is equal to or higher than zero. Using that methodology, we were able to correct the MAIAC AOD data by PBL height: we used morning PBL height estimates to correct the data retrieved from the Terra satellite and afternoon RML height estimates to correct the data retrieved from the Aqua satellite. Additional details regarding those methods can be found elsewhere, including the studies conducted by Stull (1988) and by Seidel et al. (2010).

The RH is also a key variable for the AOD–PM relationship. In the automated beta attenuation monitor employed by the CETESB (CETESB, 2017; Xavier, 2011), the PM concentrations are determined after heating the sample tube above ambient temperature to remove condensation, as well as to allow the temperature and humidity to be standardized in all measurements (Thermo Fisher Scientific Inc., 2010). Therefore, PM measured in this way are a dry aerosol mass, whereas AOD is measured under environmental conditions and is influenced by the RH. Depending on the hygroscopic nature of the aerosols, an increase in RH can modify their chemical composition, size distribution, and other characteristics; when the RH is higher, aerosol particles scatter more light, affecting the AOD loadings (Liu et al., 2013). The ratio between the aerosol scattering in an ambient air parcel with a specific RH and the aerosol scattering in dry air conditions is known as the hygroscopic growth factor— $f(RH)$ —which depends on the hygroscopic characteristics of the aerosol, the time of day, and the location (Day and Malm, 2001). Another classical correction, displayed in equation (2), accounts for the increase in aerosol loading due to an increase in RH:

$$AOD_{dry} = \frac{AOD}{f(RH)} \quad (2)$$

where AOD_{dry} is the AOD corrected by hygroscopic growth factor.

The $f(RH)$ is defined by performing specific experiments (Li et al., 2005). In the present study, we adopted the $f(RH)$ functions obtained from experiments performed by Rodrigues (2014a, b) between 2013 and 2014 in São Paulo. She fitted the $f(RH)$ function shown in equation (3) (Hänel, 1976) to the data generated by Lidar Raman techniques and found rates of growth of the aerosol under increasing RH values (γ) (standard deviation) equal to 0.11 (0.03), 0.48 (0.12) and 0.78 (0.16) on specific days in the fall, spring and summer seasons, respectively, when considering $RH_0 = 40\%$. In the period of our study, we used those values in the functions to correct the AOD retrieved in the fall, spring and summer, respectively; we used the value obtained in the fall in the function to correct the AOD retrieved in the winter.

$$f(RH) = \left(\left(1 - \frac{RH}{100} \right) / \left(1 - \frac{RH_0}{100} \right) \right)^{-\gamma} \quad (3)$$

We obtained hourly RH values from 14 of the CETESB stations and 29 monitoring stations operated by the São Paulo Municipal Climate Emergency Management Center (<https://www.cgesp.org/v3/>). For sites at which RH measurements were unavailable, the RH values obtained at the nearest station were used.

In this study, we evaluated the correlation between PM concentrations and MAIAC AOD values, considering the corrections depicted in equations (1) and (2), as well as the combination of both corrections, as

displayed in equation (4):

$$ASH_{dry} = \frac{AOD}{f(RH) \times PBLH} \quad (4)$$

where ASH_{dry} is the ASH corrected by hygroscopic growth factor.

2.4. Integration data and intercomparison

All statistical analyses were carried out with the R statistical software, version 3.5.0 (www.r-project.org). We calculated the Pearson coefficient R for the relationship between the Terra and Aqua retrievals of MAIAC AOD data and collocated PM concentrations (in the same MAIAC AOD grid cell), for all three averaging intervals during the 2012–2017 period. In our calculation of the mean diurnal and daily PM concentrations, we included only days on which at least two thirds of the measurements were valid (CETESB, 2017).

In order to eliminate atypical AOD values, which were identified in the analysis of the dispersion graphs between AODs and PM, we excluded all AOD with values above the 99th percentile from our analyses. The atypical AOD values occurred in the spring, when the MASP is impacted by long-range transport of smoke from biomass burning. This takes place in the countryside of São Paulo state, in the northern/central-west regions of Brazil, and in neighboring countries, including Bolívia, Paraguay, and Argentina, thus contributing to AOD enhancement in the MASP (Mariano et al., 2010; de Miranda et al., 2017; Yamasoe et al., 2017; Vara-Vela et al., 2018).

3. Results and discussion

3.1. Correlations between PM concentrations and MAIAC AODs

The correlations between PM concentrations and MAIAC AODs, for the satellites overpass, in the period as a whole and in each season, based on data from all of the CETESB stations, are presented in Table 1. Table S2 in the Supplementary Information (SI) presents the same information but for diurnal and daily averaging intervals. Seasons are defined as summer (December–February), fall (March–May), winter (June–August), and spring (September–November). For both PMs, the correlations were relatively weak. The strongest correlations were between the PM concentrations during Aqua overpasses and the MAIAC Aqua AODs. The MAIAC Terra AODs and the PM concentrations presented mostly negative correlations. Comparing the seasons in each averaging interval, the strongest correlations were observed in the spring.

The differences between satellites in terms of the strength of the correlations can be explained by the aerosol mixing state in the atmospheric column. The Terra satellite passes over the MASP during the mid-morning hours, when the PBL begins to evolve and, as a

Table 1

Linear correlations, by period/season, between PM concentrations and MAIAC AOD values, for the satellites overpass.

Satellite	Period/season	PM ₁₀		PM _{2.5}	
		n	R	n	R
Aqua	2012–2017	3563	0.201	1345	0.229
	Summer	230	0.236	50	0.284
	Fall	503	0.243	174	0.144
	Winter	1949	0.179	759	0.241
	Spring	881	0.249	362	0.245
Terra	2012–2017	5681	−0.031	1942	0.021
	Summer	1019	−0.119	268	−0.030
	Fall	1159	−0.030	352	0.003
	Winter	2086	−0.089	768	−0.056
	Spring	1417	0.111	554	0.120

Summer, December–February; Fall, March–May; Winter, June–August; Spring, September–November.

consequence, aerosols are not well mixed in the atmospheric column. In contrast, the Aqua satellite passes over the MASP in the early afternoon, when the PBL has evolved and aerosols are well mixed in the atmospheric column.

In addition, some studies demonstrated that the quality of the MODIS Terra retrievals worsened in MODIS Collection 5, due to the degradation of the blue band sensor. (Levy et al., 2010; Wang et al., 2012; DeVisser and Messina, 2013; Zhang et al., 2017). Even after the improvements in MODIS Collection 6, in 2014, the degradation of the MODIS Terra sensor affected MAIAC Terra AODs “via striping and a positive bias on the left-hand side of the MODIS scan, mostly over bright surfaces” (Lypustin et al., 2018).

In earlier studies, mainly conducted in the United States, weak correlations between ground-level PM_{2.5} concentrations, and AODs were reported. They found that correlations were stronger in the eastern portion of the United States ($R = 0.4$ – 0.8) than in the western portion ($R = 0.2$ – 0.4) (Toth et al., 2014; Engel-Cox et al., 2004; Paciorek et al., 2008; Zhang et al., 2009b). Loria-Salazar et al. (2016) suggested that the correlations are stronger in the eastern portion of the United States because of the well-behaved boundary layer, the more uniform vegetation coverage (which leads to lower surface reflectance), and the lower elevations, as well as the homogeneous distribution of sulphates and organic carbon.

Koelemeijer et al. (2006) reported that PM (hourly and daily) concentrations correlated weakly with MODIS (collection 4)-derived AOD in Europe. Using a decadal dataset, Hersey et al. (2015) studied the AOD–PM relationship across five major metropolitan areas in South Africa and also found weak, negative correlations between mean monthly PM concentrations and AOD.

To our knowledge, only Natali (2008) has investigated the relationship that ground-level PM concentrations have with MODIS AODs in the MASP. The author used the methodology proposed by de Almeida Castanho et al. (2008) to retrieve AODs at high resolution ($1 \times 1 \text{ km}^2$) over the MASP between 2002 and 2005, reporting correlations that were weak, negative, or both. Because there were few stations monitoring PM_{2.5} in the MASP during the 2002–2005 period, that author analyzed PM_{2.5} data from only two ground-based stations. Therefore, in the present study, we summarize the PM₁₀ results from that study. We found that the Aqua AODs correlated positively (although weakly) with the PM₁₀ concentrations during satellite overpasses ($R = 0.03$) and during the diurnal averaging intervals ($R = 0.02$), whereas the Terra AODs correlated negatively with both ($R = -0.09$ and $R = -0.1$, respectively). The correlations between the AODs and the mean PM₁₀ concentrations during the daily averaging intervals were negative for the Aqua and Terra satellites ($R = -0.06$ and $R = -0.20$, respectively). Natali (2008) identified patterns comparable to those observed in the present study.

3.2. Effect of PBL height and RH on the AOD–PM relationship

The correlations between PM concentrations and MAIAC AODs (corrected on the basis of equations (1), (2) and (4)) in the study period as a whole and for each season, for satellites overpass, based on data from all of the CETESB stations, are presented in Table 2. SI Table S3 presents the same information but for the diurnal and daily averaging intervals. Because of missing data related to PBL height estimates, the correlations in Table 1 and SI Table S2 were recalculated and are presented in Table 2 and SI Table S3, respectively, for comparison.

The fact that some PBL height (balloon sounding) data were missing was due to technical problems during balloon launches, problems during data acquisition, or problems related to the estimation method, such as when it was unable to detect inversions in the profiles (mainly during stable conditions). The mean reduction in the sample size associated with the missing PBL height estimates was 38% and 53% for the morning and afternoon periods, respectively. The frequency of missing data was higher in the afternoon period because the afternoon PBL heights were estimated indirectly on the basis of the RML height, which is not recorded every day. For the morning and afternoon periods, we attempted to use PBL height estimates from the Climate Forecast System Reanalysis, version 2, developed by the National Center for Environmental Prediction (<https://rda.ucar.edu>). However, those estimates did not present good agreement with the data obtained from soundings and were therefore not employed as a proxy for the missing data.

After recalculation of the correlations in Table 1 and SI Table S2, the pattern of the results did not change. Slight improvements were observed after the correction by PBL height (equation (1)), mainly for the MAIAC AOD Terra data, suggesting that the assumption of a well-mixed PBL is not suitable for the MASP (Tsai et al., 2011; Chu et al., 2013). Therefore, information on the vertical profiles of aerosols, provided by Lidar instruments, at ground level or aboard satellites (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation), appears crucial to the further investigation of the AOD–PM relationship. There is a need for investigations of this specific topic in the MASP.

The correlations were also improved by the f(RH) correction (equation (2)), for both satellites and all three averaging intervals. For the Aqua data, the improvements after f(RH) correction were more pronounced than were those achieved by PBL height correction, although the inverse was seen for the Terra data. In general, the improvements achieved after both corrections (equation (4)) were more pronounced than those achieved by applying only one. Despite the improvements, the correlations remained weak, for both satellites and all three intervals. In addition, the results related to the Aqua data should be interpreted with caution, because more than 50% of those data (mainly PBL height data) were missing after the corrections. For both

Table 2

Linear correlations, by period/season, between PM concentrations and MAIAC AOD values, corrected as in equation 1 (ASH), as in equation (2) (AODdry), and as in equation (4) (ASHdry), for the satellites overpass.

PM	Period/season	Aqua ^a					Terra ^b				
		n	AOD	ASH	AODdry	ASHdry	n	AOD	ASH	AODdry	ASHdry
PM ₁₀	2012–2017	1649	0.176	0.195	0.265	0.280	3480	−0.076	0.153	−0.007	−0.020
	Summer	103	0.319	0.317	0.346	0.277	623	−0.184	−0.142	−0.033	−0.021
	Fall	231	0.320	0.298	0.338	0.314	669	−0.032	0.188	0.001	0.234
	Winter	850	0.110	0.122	0.146	0.168	1087	−0.074	0.160	−0.025	0.171
	Spring	465	0.214	0.248	0.328	0.353	1101	0.014	0.176	0.033	0.099
PM _{2.5}	2012–2017	549	0.241	0.290	0.264	0.318	1014	0.034	0.064	0.236	0.025
	Summer	17	0.272	0.142	0.193	0.017	124	−0.093	0.161	−0.100	0.163
	Fall	43	0.267	0.275	0.292	0.295	117	0.073	0.159	0.318	0.399
	Winter	269	0.176	0.208	0.181	0.223	333	−0.005	0.027	0.225	0.196
	Spring	220	0.321	0.355	0.335	0.385	440	0.084	−0.019	0.230	−0.011

ASH, AOD corrected by PBL height; AOD_{dry}, AOD corrected by f(RH); ASH_{dry}, AOD corrected by PBL height and f(RH).

Summer, December–February; Fall, March–May; Winter, June–August; Spring, September–November.

^a AOD corrected by the RML height estimated by the air temperature gradient method from 00:00 UTC balloon sounding data.

^b AOD corrected by the PBL height estimated by the parcel method from 12:00 UTC balloon sounding data.

satellites and PMs, the seasonal correlations were typically stronger than were the annual correlations, although they were still weak, even after the corrections by PBL height and $f(RH)$.

In an attempt to understand why the correlations were weak or negative in the MASP, Natali (2008) focused the investigation on diurnal PM₁₀ data and found that the AODs, from the Terra and Aqua satellites, correlated more strongly (although still weak) with PM₁₀ concentrations during the winter, when the sensor scattering angle was less than 140°, and on cloud-free days. In addition, for some stations, the author evaluated the AOD–PM₁₀ relationship considering the quantity of water vapor content in the atmospheric column and concluded that high AOD values were associated with high water vapor content in the atmosphere. However, the author did not explore the effects of PBL height and $f(RH)$, as was done in the present study.

In the United States, the poor correlations between PM_{2.5} and AODs reported were justified as: the vertical distribution of the aerosols in the atmospheric column/aerosol transport at elevated levels (Toth et al., 2014; Paciorek et al., 2008), the variability in the aerosol type (Engel-Cox et al., 2006), surface reflectance (Engel-Cox et al. 2004, 2006; Paciorek et al., 2008; Zhang et al., 2009b) and the type of aerosol model used by MODIS in the AOD algorithm retrieval (Engel-Cox et al. 2004, 2006).

3.3. Evaluation of MODIS–MAIAC data over the MASP

The expected error (EE) envelope, defined as $\pm (0.05 + 0.05 \times AOD)$, is used in order to evaluate the accuracy of MAIAC AOD retrievals, and it is expected that two thirds of the retrievals (approximately 66% or one standard deviation) fall within that envelope (Martins et al., 2017). The validation study of MAIAC data for South America showed that in urban area (São Paulo and Buenos Aires), 54.7% and 57.7% of the MAIAC Terra and Aqua AOD retrievals, respectively, fell within the EE envelope. Considering the MASP data, despite the relatively strong correlations between MAIAC AODs and AERONET AODs, we found that those percentages were 38% and 45%, for the Terra and Aqua retrievals, respectively (Fig. 2). Because MODIS AOD retrievals are based mainly on the aerosol model specification and surface reflectance estimation (Levy et al., 2007; de Almeida Castanho et al., 2008; Lyapustin et al., 2011b, 2018), these results indicate that one or both are inappropriate in the MAIAC AOD retrieval algorithm.

The intercept and slope of the linear regression between AOD retrieval and AERONET AOD can be used in order to evaluate the uncertainties related to the surface reflectance estimation and aerosol model specification in the AOD retrieval algorithm, respectively. An intercept close to 0 and a slope close to 1 indicate good assumptions about the surface reflectance and the aerosol model algorithm, respectively (Sayer et al., 2013). The evaluation of those quantities for MAIAC AOD retrievals over the MASP suggested the following: the surface reflectance is better estimated during Aqua overpasses than during Terra overpasses (intercepts of -0.0003 and 0.0190 , respectively); and the

aerosol model specified by the MAIAC AOD retrieval algorithm for Aqua and Terra overpasses might not be suited for use in the MASP (slopes of 0.67 and 0.52, respectively).

The MAIAC algorithm has improved the estimation of surface spectral ratios, a central component of aerosol retrievals. Unlike the MODIS DT algorithm, which is based on prescribed spectral reflectance ratios, the MAIAC algorithm dynamically derives surface spectral ratios using the time series approach (Lyapustin et al., 2018), thus avoiding empirical assumptions about surface properties (Martins et al., 2017). Therefore, MAIAC typically increases AOD accuracy over brighter surfaces, as demonstrated in studies conducted by Lyapustin et al. (2011b) over the United States and by Zhang et al. (2019) over China.

The MAIAC algorithm has eight background aerosol models (Lyapustin et al., 2018). The background aerosol model specified by the MAIAC AOD retrieval algorithm for South America is the same as that specified for the eastern United States. That model is based on the regional climatology data obtained at the Goddard Space Flight Center (GSFC) AERONET site, in Greenbelt, Maryland, which include the aerosol characteristics of the east coast of United States (Lyapustin et al., 2018).

In Fig. 3, we present the data obtained at the GSFC and São Paulo AERONET sites in the 2000–2011 period, using the single scattering albedo (SSA) and fine-mode fraction (FMF) data to classify the aerosols by type, as proposed by Lee et al. (2010). The SSA is defined as the ratio of the scattering coefficient to the extinction coefficient of aerosols and indicates the relative capacity of the aerosols to absorb radiation at a given wavelength, a higher SSA translates to lower absorption of solar radiation (Lee et al., 2010). The FMF is defined as the ratio of fine-mode AOD (of particles with a diameter $< 1 \mu m$) to the total AOD and indicates the predominant size (fine or coarse) of the aerosols under study. According to the classification proposed by Lee et al. (2010), fine-mode aerosols are those with an FMF above 0.6 and coarse-mode aerosols are those with an FMF below 0.4. Aerosols with an FMF between 0.4 and 0.6 are designated aerosol mixtures.

As depicted in Fig. 3, the São Paulo and GSFC AERONET sites are both dominated by fine-mode aerosols. However, the aerosols evaluated at the GSFC site present a relatively homogeneous SSA, whereas those evaluated at the São Paulo site presents high variability in the degree of absorption of solar radiation. These opposing patterns were described by Levy et al. (2007), who showed that the eastern United States was dominated by non-absorbing aerosols, whereas South America (including the MASP) was dominated by aerosols that are more absorptive. Despite the differences between the two sites in terms of the aerosol properties, the aerosol model specified by the MAIAC AOD retrieval algorithm is the same for both regions. That indicates that there is a mismatch between the current MAIAC aerosol model and the local aerosol properties observed across the MASP.

The MODIS AOD products are also retrieved using the DT algorithm, at spatial resolutions of $10 \times 10 \text{ km}^2$ and $3 \times 3 \text{ km}^2$ (Levy et al., 2013), the Deep Blue algorithm (Hsu et al., 2013) and the algorithm that

Fig. 2. Comparison between the AERONET data for AOD at 550 nm (AOD₅₅₀) and those derived from the MAIAC algorithm for the Aqua satellite (left) and Terra satellite (right) in the MASP. The red and grey dashed lines indicate the MAIAC EE envelope, defined as $\pm (0.05 + 0.05 \times AOD)$ and 1:1 lines, respectively. The solid line displays the regression line. R is the linear correlation between MAIAC AODs and AERONET AODs, n is the number of data pairs (MAIAC AODs; AERONET AODs), and within EE is the fraction of MAIAC retrievals that fell within the EE envelope. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Fig. 3. The aerosol classification proposed by Lee et al. (2010), showing the dispersion of the SSA at 440 nm (SSA_{440}) and FMF, by period (season), for the AERONET sites at the GSFC (left panel) and in São Paulo (right panel), in the 2000–2011 period. FH-A, fine highly absorbing; FM-A: fine moderately absorbing; FS-A: fine slightly absorbing; FN-A: fine non-absorbing; M-A: mixed absorbing; MN-A: mixed non-absorbing; winter, June–July–August; spring, September–October–November; summer, December–January–February; fall, March–April–May.

combines the two (Levy et al., 2013), both at a spatial resolution of $10 \times 10 \text{ km}^2$. The coarse spatial resolutions of the MODIS AOD products have been an important limitation for their application in air quality and epidemiological studies at urban areas such as the MASP and could explain the lack of studies applying AOD data in this region.

A comprehensive comparison of MODIS AOD products over the MASP was beyond the scope of the present study. However, to increase our confidence in the conclusions drawn, we compared the AOD retrievals from the DT algorithm at $3 \times 3 \text{ km}^2$ (collection 6.1) with those from the AERONET site. We also calculated the validation measures—the R and the fraction of retrievals that fell within the EE envelope, as well as the slope and intercept of the regression line between DT at $3 \times 3 \text{ km}^2$ AODs and AERONET AODs—and compared them with those obtained for the MAIAC. Although the DT at $3 \times 3 \text{ km}^2$ AOD product is not suited for epidemiological applications in a relatively small region such as the MASP, we chose it for the comparison because it is the product for which the spatial resolution is closest to that of the MAIAC AOD product. In the comparison between the DT at $3 \times 3 \text{ km}^2$ AODs and the AERONET AODs, we used data for the 2012–2017 period, applying the same spatial and temporal windows used in the validation study of MAIAC data for South America ($25 \times 25 \text{ km}^2$ and 1 h, respectively). There is a need for a comprehensive validation study of the MODIS DT at $3 \times 3 \text{ km}^2$ AOD product over the MASP.

The correlations between AOD retrievals from the AERONET site and from the DT at $3 \times 3 \text{ km}^2$ algorithm were strong for the Terra and Aqua retrievals (0.83 and 0.88, respectively) and were comparable to those obtained for the MAIAC retrievals. For the DT at $3 \times 3 \text{ km}^2$ algorithm, the EE envelope was defined as $\pm (0.05 + 0.2 \times \text{AOD})$, as described by Levy et al. (2013); 86.6% and 87.0% of the Terra and Aqua retrievals, respectively, fell within that envelope. The intercept (slope) of the regression lines between AERONET AODs and DT at $3 \times 3 \text{ km}^2$ AODs were 1.0049 (0.0174) and 1.048 (0.027) for the Terra and Aqua retrievals, respectively. As expected, the DT at $3 \times 3 \text{ km}^2$ algorithm overestimated the surface reflectance at the MASP, because it is more suitable for use over dark surfaces. The results suggest that, for surface reflectance estimation, the MAIAC algorithm performs better than does the DT at $3 \times 3 \text{ km}^2$ algorithm for the Aqua retrievals and that the two algorithms have a similar performance for the Terra retrievals. For aerosol model specification, the DT at $3 \times 3 \text{ km}^2$ algorithm appears to perform better than does the MAIAC algorithm for retrievals from both satellites, because the former assumes a more absorbing aerosol optical model, which is more suitable for use over the MASP (Levy et al., 2013). Nevertheless, there are opportunities for improvements in the aerosol models for both algorithms over this region.

The atmosphere over the MASP contains a wide variety of aerosol types. Local aerosols are composed mainly of vehicle and industrial emissions. During the fire season (August to September), the MASP is

also affected by the long-range transport of smoke from biomass burning in the countryside of the state of São Paulo, in the northern/central-west regions of Brazil, and in neighboring countries, including Bolívia, Paraguay, and Argentina (Mariano et al., 2010; de Miranda et al., 2017; Yamasoe et al., 2017; Vara-Vela et al., 2018). The studies conducted by Martins et al. (2009) and Yamasoe et al. (2017) showed that the local aerosols are more absorptive than are those transported from areas of biomass burning. These results explain the high AOD variability, mainly in the spring (Fig. 3). Furthermore, the MASP is affected by sea-breeze circulation (Freitas et al., 2007), an additional aerosol source that also contributes to the high aerosol variability in the MASP. All of those aerosol sources increase the variability in aerosol properties, making AOD retrieval more complex (Engel-Cox et al., 2006; Lee et al., 2016).

The RH effect on the AOD–PM relationship can also be related to the aerosol model specification in the MAIAC AOD retrieval algorithm, because RH can change the chemical and extinction characteristics of aerosols (personal communication with Alexei Lyapustin, 2019). According to the fraction of hygroscopic components present in the aerosol, the increase in RH can enhance the aerosol extinction, due to the increase in the scattering efficiency (Liu et al., 2013). Since sulphates and nitrates, which are highly hygroscopic, are the main species in the aerosol in São Paulo (Pereira et al., 2017) and in the Eastern United States (Bell et al., 2007), it is possible that an increase in the RH change the SSA of aerosols in the MASP, and make it comparable to that of aerosols in the eastern United States. That would explain why MAIAC uses the same aerosol model for both locations, despite the differences in the SSA properties shown in Fig. 3.

To elucidate that, Fig. 4 presents the mean SSA for each bin of FMF and Total Precipitable Water (TPW) at the GSFC and São Paulo AERONET sites. Changes in TPW are caused in part by RH and in part by seasonal variation, because TPW is a function of temperature and RH (Lee et al., 2010). The increase caused by a change in SSA for TPW was lower in São Paulo than at GSFC. Therefore, even if the RH effect is taken into account, the aerosols in São Paulo appear to be more absorptive than those in the eastern United States, highlighting the mismatch between the current MAIAC aerosol model and the aerosol properties observed across the MASP.

Although in both locations sulphates and nitrates are the dominant species, which are highly reflective of solar radiation (Wallace and Hobs, 2006), and explain the high SSA in both locations, the relative amount of elemental carbon (EC)/black carbon (BC) in the $PM_{2.5}$ mass of aerosol in São Paulo is at least four times higher than that in the Eastern US. The study conducted by Pereira et al. (2017) in São Paulo, reported an EC contribution of about 16% to the $PM_{2.5}$ mass, while the contribution of the EC to the $PM_{2.5}$ mass in the Eastern US reported by Bell et al. (2007) was about 4%. The studies conducted by de Almeida Castanho and Artaxo (2001) and Andrade et al. (2012b) in São Paulo reported a BC

Fig. 4. Mean SSA at 440 nm (SSA_{440}) for each bin of FMF and TPW for the AERONET sites at the GSFC (left panel) and in São Paulo (right panel). Only bins with more than five data points are shown.

contribution of about 25% and 36% to the $PM_{2.5}$ mass, respectively. On the other hand, the study by [de Almeida Castanho et al. \(2005\)](#) reported a BC contribution of 3% of the $PM_{2.5}$ mass on the East Coast of the United States. Since EC/BC are, in general, strong absorbing, SSA in São Paulo is lower than the aerosol in the Eastern US.

In previous versions of the MODIS AOD retrieval algorithms, aerosol models were updated to improve the representativeness of the physical and optical properties of local aerosols. In MODIS collection 5, a seasonal (winter and spring) absorbing aerosol model with an SSA of approximately 0.85 was assigned to the region comprising the south-central and southeastern regions of Brazil, whereas a moderately absorbing aerosol model (SSA of approximately 0.90) was assigned to the remaining regions and for the summer and fall ([Levy et al., 2007](#), [Fig. 3](#)). The MAIAC AOD is a relatively new product for which updates are underway. In the upcoming version of the MAIAC algorithm, retrievals will take into account the effects of regional aerosol seasonality, such as those caused by variations in humidity or by periods of biomass burning ([Lyapustin et al., 2018](#)). Until then, the MAIAC AOD data should not be used as a predictor of ground-level PM concentrations in the MASP.

The results of the present study indicate that MAIAC AOD retrievals over this region are affected mainly by the mismatch between the current MAIAC aerosol model for the MASP and the local aerosol properties. Compared with the DT algorithm at $3 \times 3 \text{ km}^2$ (collection 6.1), the MAIAC surface reflectance estimation is more accurate for Aqua retrievals than for Terra retrievals.

4. Conclusions

This study examined the high-resolution AODs derived from the MAIAC algorithm as a predictor of ground-level PM concentrations at the MASP. We calculated the correlations between ground-level PM concentrations (at the hour of the satellites overpass, diurnal averaging, and daily averaging intervals) and the collocated AODs. We found those correlations to be weak. Our findings corroborate those of the study conducted by [Natali \(2008\)](#). Although the correlations were slightly improved after corrections by PBL height and $f(RH)$, they remained weak. Our results indicate that the mismatch between the current aerosol model specified by the MAIAC algorithm for the MASP and the physical and optical properties of the local aerosols is the main issue affecting MAIAC AOD retrievals. Therefore, high-resolution AODs from the current MAIAC version will not serve well to predict ground-level PM concentrations in the MASP.

The MAIAC algorithm would greatly benefit from an improved aerosol model applied to MASP. Our results suggest that sources of aerosol particles in the MASP are quite diverse and that there is no single

optical model suitable for use with satellite-derived AODs in the region. In this context, [de Almeida Castanho et al. \(2008\)](#) presented a tool that, using critical reflectance, dynamically assigns an SSA value to the aerosol model that reflects the optical properties of the aerosol layer in the MASP. For the spectral reflectance ratio, those authors used the prescribed ratio of 0.6. Therefore, we recommend that future studies employ an algorithm combining the dynamic specification of surface spectral ratios from the MAIAC algorithm and the dynamic specification of SSA in the aerosol model proposed by [de Almeida Castanho et al., \(2008\)](#). We believe that such a combination would improve AOD retrievals over the MASP, because it will combine the improvements of the MAIAC algorithm over bright surfaces and the correct specification of the aerosol model, taking advantage of its higher spatial resolution.

As seen in applications in other regions, high resolution MAIAC AOD retrievals could be assimilated into statistical (e.g., [Chudnovsky et al., 2014b](#)), geophysical (e.g., [Jin et al., 2019](#)) or a combination of both approaches (e.g., [Meng et al., 2019](#)), in order to retrospectively predict PM concentrations, resulting in PM databases for future use in health studies in the MASP.

CRedit authorship contribution statement

Aline Santos Damascena: Conceptualization, Software, Formal analysis, Investigation, Writing - original draft. **Márcia Akemi Yamasoe:** Conceptualization, Supervision, Writing - review & editing. **Vitor Souza Martins:** Investigation, Software, Validation. **Jorge Rosas:** Investigation, Software. **Noelia Rojas Benavente:** Investigation, Software. **Maciel Piñero Sánchez:** Investigation, Software. **Nelson Ithiro Tanaka:** Conceptualization, Supervision. **Paulo Hilário Nascimento Saldiva:** Conceptualization, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors are grateful to the CETESB, to the Earthdata team, and to the São Paulo Municipal Climate Emergency Management Center, for providing the air quality data, MAIAC AOD data, and meteorological data, respectively. We thank the AERONET principal investigators, Paulo Artaxo and Brent Holben, for providing sun photometric data from the São Paulo and GSFC sites, respectively. Finally, we greatly appreciate the anonymous reviewers for their valuable comments, which have

significantly improved the manuscript. This study is also part of the MetroHealth subproject of a project entitled “The Use of Modern Autopsy Techniques to Investigate Human Diseases (MODAU)”, funded by the *Fundação de Amparo à Pesquisa do Estado de São Paulo* (FAPESP, São Paulo Research Foundation, Brazil). Aline Santos Damascena is the recipient of a PhD scholarship from FAPESP (grant no. 2016/09411-1) and of a PhD scholarship from the Brazilian *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (Capes, Office for the Advancement of Higher Education) (grant no. 001).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.atmosenv.2020.117949>.

References

- Albuquerque, T.T.A., Andrade, M.F., Ynoue, R.Y., 2012. Characterization of atmospheric aerosols in the city of São Paulo, Brazil: comparisons between polluted and unpolluted periods. *Environ. Monit. Assess.* 184 (2), 969–984.
- de Almeida Castanho, A.D., Artaxo, P., 2001. Wintertime and summertime São Paulo aerosol source apportionment study. *Atmos. Environ.* 35 (29), 4889–4902.
- de Almeida Castanho, A.D., Artaxo, P., Martins, J.V., Hobbs, P.V., Remer, L., Yamasoe, M., Colarco, P.R., 2005. Chemical characterization of aerosols on the East Coast of the United States using aircraft and ground-based stations during the CLAMS experiment. *J. Atmos. Sci.* 62 (4), 934–946.
- de Almeida Castanho, A.D., Martins, J.V., Artaxo, P., 2008. MODIS aerosol optical depth retrievals with high spatial resolution over an urban area using the critical reflectance. *J. Geophys. Res.: Atmosphere* 113, D02201. <https://doi.org/10.1029/2007JD008751>.
- Alves, E., Rodrigues, C., Braga, A., Miranda, M., Saldiva, P.H.N., 2014. Monitoring Air Quality in Brazil. São Paulo, Instituto Saúde e Sustentabilidade (in Portuguese). <http://www.saudeesustentabilidade.org.br/site/wp-content/uploads/2014/07/Monitoramento-da-Qualidade-do-Ar-no-Brasil-2014.pdf>.
- Andrade, M.F., Fornaro, A., Mazzoli, C.R., Martins, L.D., Boian, C., Oliveira, M.G.L., Carbone, S., Peres, J., Alvalá, P., Leme, N.P., 2012a. Ozone sounding in the metropolitan area of São Paulo, Brazil: wet and dry season campaigns of 2006. *Atmos. Environ.* 61, 627–640.
- Andrade, M.F., de Miranda, R.M., Fornaro, A., Kerr, A., Oyama, B., de Andre, P.A., Saldiva, P., 2012b. Vehicle emissions and PM2.5 mass concentrations in six Brazilian cities. *Air Qual. Atmos. Health* 5 (1), 79–88.
- Bell, M.L., Dominici, F., Ebisu, K., Zeger, S.L., Samet, J.M., 2007. Spatial and temporal variation in PM2.5 chemical composition in the United States for health effects studies. *Environ. Health Perspect.* 115 (7), 989–995.
- Beloconi, A., Chrysoulakis, N., Lyapustin, A., Utzinger, J., Vounatsou, P., 2018. Bayesian geostatistical modelling of PM10 and PM2.5 surface level concentrations in Europe using high-resolution satellite-derived products. *Environ. Int.* 121, 57–70.
- Boyounk, N., Léon, J.F., Delbarre, H., Podvin, T., Deroo, C., 2010. Impact of the mixing boundary layer on the relationship between PM2.5 and aerosol optical thickness. *Atmos. Environ.* 44 (2), 271–277.
- Braga, A.L., Saldiva, P.H.N., Pereira, L.A., Menezes, J.J., Conceição, G.M., Lin, C.A., Zanobetti, A., Schwartz, J., Dockery, D.W., 2001. Health effects of air pollution exposure on children and adolescents in São Paulo, Brazil. *Pediatr. Pulmonol.* 31 (2), 106–113.
- Carvalho, V.S.B., Freitas, E.D., Martins, L.D., Martins, J.A., Mazzoli, C.R., Andrade, M.F., 2015. Air quality status and trends over the Metropolitan Area of São Paulo, Brazil as a result of emission control policies. *Environ. Sci. Pol.* 47, 68–79.
- CETESB, 2016. Classification of Spatial Representativeness of Air Quality Monitoring Stations in the State of São Paulo third step - Report (in Portuguese). <https://cetesb.sp.gov.br/ar/publicacoes-relatorios/>.
- CETESB, 2017. Air Quality in São Paulo State Report (in Portuguese). <https://cetesb.sp.gov.br/ar/publicacoes-relatorios/>.
- Chew, B.N., Campbell, J.R., Hyer, E.J., Salinas, S.V., Reid, J.S., Welton, E.J., Holben, B.N., Liew, S.C., 2016. Relationship between aerosol optical depth and particulate matter over Singapore: effects of aerosol vertical distributions. *Aerosol Air Qual. Res.* 16, 2818–2830.
- Chu, D.A., Kaufman, Y.J., Zibordi, G., Chern, J.D., Mao, J., Li, C., Holben, B.N., 2003. Global monitoring of air pollution over land from the earth observing system-terra moderate resolution imaging spectroradiometer (MODIS). *J. Geophys. Res.: Atmosphere* 108, D21. <https://doi.org/10.1029/2002JD003179>.
- Chu, D.A., Tsai, T.C., Chen, J.P., Chang, S.C., Jeng, Y.J., Chiang, W.L., Lin, N.H., 2013. Interpreting aerosol lidar profiles to better estimate surface PM2.5 for columnar AOD measurements. *Atmos. Environ.* 79, 172–187.
- Chu, Y., Liu, Y., Li, X., Liu, Z., Lu, H., Lu, Y., Mao, Z., Chen, X., Li, N., Meng, R., Liu, F., Tian, L., Zhu, Z., Xiang, H., 2016. A review on predicting ground PM2.5 concentration using satellite aerosol optical depth. *Atmosphere* 7 (10), 129.
- Chudnovsky, A.A., Lyapustin, A., Wang, Y., Tang, C., Schwartz, J., Koutrakis, P., 2014a. High resolution aerosol data from MODIS satellite for urban air quality studies. *Open Geosci.* 6 (1), 17–26.
- Chudnovsky, A.A., Koutrakis, P., Kloog, I., Melly, S., Nordio, F., Lyapustin, A., Schwartz, J., 2014b. Fine particulate matter predictions using high resolution Aerosol Optical Depth (AOD) retrievals. *Atmos. Environ.* 89, 189–198.
- Day, D.E., Malm, W.C., 2001. Aerosol light scattering measurements as a function of relative humidity: a comparison between measurements made at three different sites. *Atmos. Environ.* 35 (30), 5169–5176.
- DeVisser, M.H., Messina, J.P., 2013. Exploration of sensor comparability: a case study of composite MODIS Aqua and Terra data. *Rem. Sens. Lett.* 4 (6), 599–608.
- van Donkelaar, A., Martin, R.V., Li, C., Burnett, R.T., 2019. Regional estimates of chemical composition of fine particulate matter using a combined geoscience-statistical method with information from satellites, models, and monitors. *Environ. Sci. Technol.* 53 (5), 2595–2611.
- Engel-Cox, J.A., Holloman, C.H., Coutant, B.W., Hoff, R.M., 2004. Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. *Atmos. Environ.* 38 (16), 2495–2509.
- Engel-Cox, J.A., Hoff, R.M., Rogers, R., Dimmick, F., Rush, A.C., Szykman, J.J., Zell, E.R., 2006. Integrating lidar and satellite optical depth with ambient monitoring for 3-dimensional particulate characterization. *Atmos. Environ.* 40 (40), 8056–8067.
- Fajersztajn, L., Veras, M., Barrozo, L.V., Saldiva, P.H.N., 2014. Air monitoring coverage in low-income countries: an observational study. *Lancet* 384, S14.
- Freitas, E.D., Rozoff, C.M., Cotton, W.R., Dias, P.L.S., 2007. Interactions of an urban heat island and sea-breeze circulations during winter over the metropolitan area of São Paulo Brazil. *Boundary-Layer Meteorol.* 122 (1), 43–65.
- Gong, W., Huang, Y., Zhang, T., Zhu, Z., Ji, Y., Xiang, H., 2017. Impact and suggestion of column-to-surface vertical correction scheme on the relationship between satellite AOD and ground-level PM2.5 in China. *Rem. Sens.* 9 (10), 1038.
- Guinot, B., Roger, J.C., Cachier, H., Pucai, W., Jianhui, B., Tong, Y., 2006. Impact of vertical atmospheric structure on Beijing aerosol distribution. *Atmos. Environ.* 40 (27), 5167–5180.
- Gupta, P., Christopher, S.A., Wang, J., Gehrig, R., Lee, Y.C., Kumar, N., 2006. Satellite remote sensing of particulate matter and air quality assessment over global cities. *Atmos. Environ.* 40 (30), 5880–5892.
- Habermann, M., Gouveia, N., 2012. Aplicação de regressão baseada no uso do solo para prever a concentração de material particulado inalável no município de São Paulo Brasil. *Eng. Sanitária Ambient.* 17, 155–162.
- Hänel, G., 1976. The properties of atmospheric aerosol particles as functions of the relative humidity at thermodynamic equilibrium with the surrounding moist air. *Adv. Geophys.* 19, 73–188.
- Hersey, S.P., Garland, R.M., Crosbie, E., Shingler, T., Sorooshian, A., Piketh, S., Burger, R., 2015. An overview of regional and local characteristics of aerosols in South Africa using satellite, ground, and modeling data. *Atmos. Chem. Phys.* 15, 4259.
- Holben, B.N., Eck, T.F., Slutsker, I., Tanré, D., Buis, J.P., Setzer, A., Vermote, E., Reagan, J.A., Kaufman, Y.J., Lavenue, F., Jankowiak, I., Smirnov, A., 1998. Aeronet - a federated instrument network and data archive for aerosol characterization. *Rem. Sens. Environ.* 66 (1), 1–16.
- Hsu, N.C., Jeong, M.J., Bettenhausen, C., Sayer, A.M., Hansell, R., Seftor, C.S., Huang, J., Tsay, S.C., 2013. Enhanced Deep Blue aerosol retrieval algorithm: the second generation. *J. Geophys. Res.: Atmosphere* 118 (16), 9296–9315. <https://doi.org/10.1002/jgrd.50712>.
- IEMA, 2014. First Diagnosis of Air Quality Monitoring Networks in Brazil. Brazil, Instituto de Energia e Meio Ambiente (in Portuguese). https://www.mma.gov.br/images/arquivo/80060/Diagnostico_Ne_de_Monitoramento_da_Qualidade_do_Ar.pdf.
- IPCC, 2013. Climate Change 2013: the physical science basis. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, p. 1535.
- Jin, X., Fiore, A.M., Curci, G., Lyapustin, A., Civerolo, K., Ku, M., van Donkelaar, A., Martin, R.V., 2019. Assessing uncertainties of a geophysical approach to estimate surface fine particulate matter distributions from satellite-observed aerosol optical depth. *Atmos. Chem. Phys.* 19 (1).
- Knibbs, L.D., van Donkelaar, A., Martin, R.V., Bechle, M.J., Brauer, M., Cohen, D.D., Cowie, C.T., Dirgawati, M., Guo, Y., Hanigan, I.C., Johnston, F.H., Marks, G.B., Marshall, J.D., Pereira, G., Jalaludin, B., Heyworth, J.S., Morgan, G.G., Barnett, A.G., 2018. Satellite-based land-use regression for continental-scale long-term ambient PM2.5 exposure assessment in Australia. *Environ. Sci. Technol.* 52 (21), 12445–12455.
- Koelemeijer, R.B.A., Homan, C.D., Matthijsen, J., 2006. Comparison of spatial and temporal variations of aerosol optical thickness and particulate matter over Europe. *Atmos. Environ.* 40 (27), 5304–5315.
- Kong, L., Xin, J., Zhang, W., Wang, Y., 2016. The empirical correlations between PM2.5 PM10 and AOD in the Beijing metropolitan region and the PM2.5 PM10 distributions retrieved by MODIS. *Environ. Pollut.* 216, 350–360.
- Kumar, A., Jiménez, R., Belalcázar, L.C., Rojas, N.Y., 2016. Application of WRF-Chem model to simulate PM10 concentration over Bogotá. *Aero. Air Qual. Res.* 16 (5), 1206–1221.
- Lee, J., Kim, J., Song, C.H., Kim, S.B., Chun, Y., Sohn, B.J., Holben, B.N., 2010. Characteristics of aerosol types from AERONET sunphotometer measurements. *Atmos. Environ.* 44 (26), 3110–3117.
- Lee, H.J., Chatfield, R.B., Strawa, A.W., 2016. Enhancing the applicability of satellite remote sensing for PM2.5 estimation using MODIS deep blue AOD and land use regression in California, United States. *Environ. Sci. Technol.* 50 (12), 6546–6555.
- Levy, R.C., Remer, L.A., Dubovik, O., 2007. Global aerosol optical properties and application to Moderate Resolution Imaging Spectroradiometer aerosol retrieval

- over land. *J. Geophys. Res.: Atmosphere* 112, D13. <https://doi.org/10.1029/2006JD007815>.
- Levy, R.C., Remer, L.A., Kleidman, R.G., Mattoo, S., Ichoku, C., Kahn, R., Eck, T.F., 2010. Global evaluation of the Collection 5 MODIS dark-target aerosol products over land. *Atmos. Chem. Phys.* 10 (21), 10399.
- Levy, R.C., Mattoo, S., Munchak, L.A., Remer, L.A., Sayer, A.M., Patadia, F., Hsu, N.C., 2013. The Collection 6 MODIS aerosol products over land and ocean. *Atmos. Measur. Tech.* 6 (11), 2989.
- Li, C., Mao, J., Lau, A.K., Yuan, Z., Wang, M., Liu, X., 2005. Application of MODIS satellite products to the air pollution research in Beijing. *Sci. China Earth Sci.* 48, 209–219.
- Li, J., Carlson, B.E., Laci, A.A., 2015. How well do satellite AOD observations represent the spatial and temporal variability of PM_{2.5} concentration for the United States? *Atmos. Environ.* 102, 260–273.
- Liu, X., Gu, J., Li, Y., Cheng, Y., Qu, Y., Han, T., Wang, J., Tian, H., Chen, J., Zhang, Y., 2013. Increase of aerosol scattering by hygroscopic growth: Observation, modeling, and implications on visibility. *Atmos. Res.* 132, 91–101.
- Loria-Salazar, S.M., Holmes, H.A., Arnott, W.P., Barnard, J.C., Moosmüller, H., 2016. Evaluation of MODIS columnar aerosol retrievals using AERONET in semi-arid Nevada and California, USA, during the summer of 2012. *Atmos. Environ.* 144, 345–360.
- Lyapustin, A., Martonchik, J., Wang, Y., Laszlo, I., Korkin, S., 2011a. Multiangle implementation of atmospheric correction (MAIAC): 1 Radiative transfer basis and look-up tables. *J. Geophys. Res.: Atmosphere* 116, D3. <https://doi.org/10.1029/2010JD014985>.
- Lyapustin, A., Wang, Y., Laszlo, I., Kahn, R., Korkin, S., Remer, L., Reid, J.S., 2011b. Multiangle implementation of atmospheric correction (MAIAC): 2 Aerosol algorithm. *J. Geophys. Res.: Atmosphere* 116, D3. <https://doi.org/10.1029/2010JD014986>.
- Lyapustin, A., Wang, Y., Laszlo, I., Hilker, T., Hall, F.G., Sellers, P.J., Korkin, S.V., 2012. Multi-angle implementation of atmospheric correction for MODIS (MAIAC): 3 Atmospheric correction. *Rem. Sens. Environ.* 127, 385–393.
- Lyapustin, A., Wang, Y., Korkin, S., Huang, D., 2018. MODIS Collection 6 MAIAC algorithm. *Atmos. Measur. Tech.* 11 (10).
- Mariano, G.L., Lopes, F.J.S., Jorge, M.P.P.M., Landulfo, E., 2010. Assessment of biomass burnings activity with the synergy of sunphotometric and LIDAR measurements in São Paulo, Brazil. *Atmos. Res.* 98 (2–4), 486–499.
- Martins, J.V., Artaxo, P., Kaufman, Y.J., de Almeida Castanho, A.D., Remer, L.A., 2009. Spectral absorption properties of aerosol particles from 350–2500nm. *Geophys. Res. Lett.* 36 (13) <https://doi.org/10.1029/2009GL037435>.
- Martins, V.S., Lyapustin, A., de Carvalho, L.A., Barbosa, C.C., Novo, E.M., 2017. Validation of high-resolution MAIAC aerosol product over South America. *J. Geophys. Res.: Atmosphere* 122 (14), 7537–7559.
- Meng, J., Li, C., Martin, R.V., van Donkelaar, A., Hystad, P., Brauer, M., 2019. Estimated long-term (1981–2016) concentrations of ambient fine particulate matter across North America from chemical transport modeling, satellite remote sensing, and ground-based measurements. *Environ. Sci. Technol.* 53 (9), 5071–5079.
- de Miranda, R.M., Andrade, M.F., Fornaro, A., Astolfo, R., de Andre, P.A., Saldiva, P.H.N., 2012. Urban air pollution: a representative survey of PM_{2.5} mass concentrations in six Brazilian cities. *Air Qual. Atmos. Health* 5 (1), 63–77.
- de Miranda, R.M., Lopes, F., Do Rosário, N.E., Yamasoe, M.A., Landulfo, E., Andrade, M. F., 2017. The relationship between aerosol particles chemical composition and optical properties to identify the biomass burning contribution to the particles concentration: a case study for São Paulo city, Brazil. *Environ. Monit. Assess.* 189 (1), 6.
- Natali, L., 2008. The Use of Remote Sensing Products to Characterize Air Quality in São Paulo Metropolitan Region. Dissertation, University of São Paulo (in Portuguese). <http://teses.usp.br/>.
- Paciorek, C.J., Liu, Y., Moreno-Macias, H., Kondragunta, S., 2008. Spatiotemporal associations between GOES aerosol optical depth retrievals and ground-level PM_{2.5}. *Environ. Sci. Technol.* 42 (15), 5800–5806.
- Pereira, G.M., Teinilä, K., Custódio, D., Gomes Santos, A., Xian, H., Hillamo, R., Alves, C. A., de Andrade, J.B., da Rocha, G.O., Kumar, P., Balasubramanian, R., Andrade, M. F., Vasconcelos, P.C., 2017. Particulate pollutants in the Brazilian city of São Paulo: 1-year investigation for the chemical composition and source apportionment. *Atmos. Chem. Phys. Discuss.* 17, 11943–11969.
- Ribeiro, H., Cardoso, M.R.A., 2003. Air pollution and children's health in São Paulo (1986–1998). *Soc. Sci. Med.* 57 (11), 2013–2022.
- Ribeiro, A.G., Downward, G.S., de Freitas, C.U., Neto, F.C., Cardoso, M.R.A., de Oliveira, M.D.R.D., Nardocci, A.C., 2019. Incidence and mortality for respiratory cancer and traffic-related air pollution in São Paulo, Brazil. *Environ. Res.* 170, 243–251.
- Rodrigues, P.F., 2014. Hygroscopicity Evaluation of Urban Aerosols by the Lidar Raman Technique. Energy and Nuclear Research Institute, University of São Paulo. PhD thesis. <http://pelicano.ipen.br/PosG30/TextoCompleto/Patricia%20Ferrini%20Rodrigues.D.pdf>.
- Rodrigues, P.F., Landulfo, E., Lopes, F.J.S., da Costa, R.F., Granados-Muñoz, M.J., Guerrero-Rascado, J.L., 2014. Evaluation of the Hygroscopic Behavior of Aerosols over São Paulo: One-Day Case Study. In *Lidar Technologies, Techniques, and Measurements for Atmospheric Remote Sensing X*. International Society for Optics and Photonics, Bellingham, WA, USA.
- Saldiva, P.H.N., Lichtenfels, A.J.F.C., Paiva, P.S.O., Barone, I.A., Martins, M.A., Massad, E., Pereira, J.C.R., Xavier, V.P., Singer, J.M., Bohm, G.M., 1994. Association between air pollution and mortality due to respiratory diseases in children in São Paulo, Brazil: a preliminary report. *Environ. Res.* 65 (2), 218–225.
- Saldiva, P.H.N., Pope III, C.A., Schwartz, J., Dockery, D.W., Lichtenfels, A.J., Salge, J.M., Barone, I., Bohm, G.M., 1995. Air pollution and mortality in elderly people: a time-series study in São Paulo, Brazil. *Arch. Environ. Health* 50 (2), 159–163.
- Sánchez, M.P., de Oliveira, A.P., Varona, R.P., Tito, J.V., Codato, G., Ribeiro, F.N.D., Filho, E.P.M., da Silveira, L.C., 2020. Rawinsonde-based analysis of the urban boundary layer in the metropolitan region of São Paulo, Brazil. *Earth Space Sci.* 7 (2).
- Sayer, A.M., Hsu, N.C., Bettenhausen, C., Jeong, M.J., 2013. Validation and uncertainty estimates for MODIS Collection 6 “Deep Blue” aerosol data. *J. Geophys. Res.: Atmosphere* 118 (14), 7864–7872.
- Schaap, M., Apituley, A., Timmermans, R.M.A., Koelemeijer, R.B.A., Leeuw, G.D., 2009. Exploring the relation between aerosol optical depth and PM_{2.5} at Cabauw, The Netherlands. *Atmos. Chem. Phys.* 9 (3), 909–925.
- Seidel, D.J., Ao, C.O., Li, K., 2010. Estimating climatological planetary boundary layer heights from radiosonde observations: comparison of methods and uncertainty analysis. *J. Geophys. Res.: Atmosphere* 115, D16113. <https://doi.org/10.1029/2009JD013680>.
- Seinfeld, J.H., Pandis, S.N., 2016. *Atmospheric Chemistry and Physics: from Air Pollution to Climate Change*. John Wiley & Sons.
- Slater, J.F., Dibb, J.E., 2004. Relationships between surface and column aerosol radiative properties and air mass transport at a rural New England site. *J. Geophys. Res.: Atmosphere* 109.
- Stafoggia, M., Schwartz, J., Badaloni, C., Bellander, T., Alessandrini, E., Cattani, G., Donato, F., Gaeta, A., Leone, G., Lyapustin, A., Sorek-Hamer, M., de Hoogh, K., Di, Q., Forastiere, F., Kloog, I., 2017. Estimation of daily PM₁₀ concentrations in Italy (2006–2012) using finely resolved satellite data, land use variables and meteorology. *Environ. Int.* 99, 234–244.
- Stull, R.B., 1988. *An Introduction to Boundary Layer Meteorology*. Kluwer Academic Publishers, Boston.
- Takano, A.P.C., Justo, L.T., dos Santos, N.V., Marquezini, M.V., de André, P.A., da Rocha, F.M.M., Pasqualucci, C.A., Barrozo, L.V., Singer, J.M., de André, C.D.S., Veras, M.M., Saldiva, P.H.N., 2019. Pleural anthracosis as an indicator of lifetime exposure to urban air pollution: an autopsy-based study in São Paulo. *Environ. Res.* 173, 23–32.
- Thermo Fisher Scientific Inc, 2010. FH62C14 Instruction Manual. <https://assets.thermo.com/Assets/Assets/Null/Package-Inserts/EPM-manual-FH62C14.pdf>.
- Toth, T.D., Zhang, J., Campbell, J.R., Hyer, E.J., Reid, J.S., Shi, Y., Westphal, D.L., 2014. Impact of data quality and surface-to-column representativeness on the PM_{2.5}/satellite AOD relationship for the contiguous United States. *Atmos. Chem. Phys.* 14 (12), 6049–6062.
- Tsai, T.C., Jeng, Y.J., Chu, D.A., Chen, J.P., Chang, S.C., 2011. Analysis of the relationship between MODIS aerosol optical depth and particulate matter from 2006 to 2008. *Atmos. Environ.* 45 (27), 4777–4788.
- Vara-Vela, A., Andrade, M.F., Zhang, Y., Kumar, P., Ynoue, R.Y., Souto-Oliveira, C.E., Lopes, F.J.S., Landulfo, E., 2018. Modeling of atmospheric aerosol properties in the São Paulo metropolitan area: impact of biomass burning. *J. Geophys. Res.: Atmosphere* 123 (17), 9935–9956.
- Wallace, J.M., Hobbs, P.V., 2006. *Atmospheric Science: an Introductory Survey*, second ed. Elsevier.
- Wang, D., Morton, D., Masek, J., Wu, A., Nagol, J., Xiong, X., Levy, R., Vermote, E., Wolfe, R., 2012. Impact of sensor degradation on the MODIS NDVI time series. *Rem. Sens. Environ.* 119, 55–61.
- World Health Organization, 2006. Regional Office for Europe. Air Quality Guidelines: Global Update 2005: Particulate Matter, Ozone, Nitrogen Dioxide, and Sulfur Dioxide. World Health Organization Europe, Copenhagen, Denmark.
- World Health Organization, 2016. Ambient Air Pollution: A Global Assessment of Exposure and Burden of Disease 2016. World Health Organization Europe, Copenhagen, Denmark.
- World Urbanization Prospects, 2019a. The 2019 Revision. United Nations Department of Economic and Social Affairs. Population Division, New York. <https://www.worldometers.info/demographics/world-demographics/#urb/>.
- World Urbanization Prospects, 2019b. The 2019 Revision. United Nations Department of Economic and Social Affairs. Population Division, New York. <https://www.worldometers.info/world-population/south-america-population/>.
- Xavier, J.C.M., 2011. Analysis of the Availability of the Automatic Air Quality Monitoring Network and its Effects on Environmental Licensing in São Paulo. Dissertation. Technological Institute of Aeronautics (in Portuguese). <https://cetesb.sp.gov.br/escolasuperior/wp-content/uploads/sites/30/2016/06/jose.xavier.pdf>.
- Yamasoe, M.A., do Rosário, N.M.E., Barros, K.M., 2017. Downward solar global irradiance at the surface in São Paulo city - the climatological effects of aerosol and clouds. *J. Geophys. Res.: Atmosphere* 122 (1), 391–404.
- Zhang, Q., Ma, X., Tie, X., Huang, M., Zhao, C., 2009a. Vertical distributions of aerosols under different weather conditions: analysis of in-situ aircraft measurements in Beijing, China. *Atmos. Environ.* 43 (34), 5526–5535.
- Zhang, H., Hoff, R.M., Engel-Cox, J.A., 2009b. The relation between Moderate Resolution Imaging Spectroradiometer (MODIS) aerosol optical depth and PM_{2.5} over the United States: a geographical comparison by US Environmental Protection Agency regions. *J. Air Waste Manag. Assoc.* 59 (11), 1358–1369.
- Zhang, Y., Song, C., Band, L.E., Sun, G., Li, J., 2017. Reanalysis of global terrestrial vegetation trends from MODIS products: browning or greening? *Rem. Sens. Environ.* 191, 145–155.
- Zhang, Z., Wu, W., Fan, M., Wei, J., Tan, Y., Wang, Q., 2019. Evaluation of MAIAC aerosol retrievals over China. *Atmos. Environ.* 202, 8–16.
- Zheng, C., Zhao, C., Zhu, Y., Wang, Y., Shi, X., Wu, X., Chen, T., Wu, F., Qiu, Y., 2017. Analysis of influential factors for the relationship between PM_{2.5} and AOD in Beijing. *Atmos. Chem. Phys.* 17 (21), 13473–13489.

2. Parecer do Comitê de Ética e Pesquisa

FACULDADE DE MEDICINA DA
UNIVERSIDADE DE SÃO
PAULO - FMUSP

PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Modelo espaço-temporal para a predição da concentração de material particulado fino na Região Metropolitana de São Paulo

Pesquisador: Paulo Hilário Nascimento Saldiva

Área Temática:

Versão: 1

CAAE: 54503816.1.0000.0065

Instituição Proponente: Faculdade de Medicina da Universidade de São Paulo

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 1.482.675

Apresentação do Projeto:

Trata-se de projeto de pesquisa meteorológico da contaminação ambiental por medidas estimadas por geoprocessamento e por medidas físicas em cascas de árvores.

Objetivo da Pesquisa:

Modelo espaço-temporal para a predição da concentração de material particulado fino na Região Metropolitana de São Paulo

Avaliação dos Riscos e Benefícios:

Projeto sem aspectos éticos humanos pertinentes.

Comentários e Considerações sobre a Pesquisa:

Adequada

Considerações sobre os Termos de apresentação obrigatória:

Não envolve intervenção em seres humanos ou animais

Recomendações:

Nenhuma

Conclusões ou Pendências e Lista de Inadequações:

Nenhuma

FACULDADE DE MEDICINA DA
UNIVERSIDADE DE SÃO
PAULO - FMUSP

Continuação do Parecer: 1.482.675

Considerações Finais a critério do CEP:

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_657158.pdf	22/03/2016 11:49:48		Aceito
Outros	FolhaRostoCadastroAssinada.pdf	22/03/2016 11:47:58	ALINE SANTOS DAMASCENA	Aceito
Folha de Rosto	FolhaRostoBrasilAssinada.pdf	22/03/2016 11:45:08	ALINE SANTOS DAMASCENA	Aceito
Declaração de Instituição e Infraestrutura	Carta_IME.pdf	29/01/2016 09:12:46	ALINE SANTOS DAMASCENA	Aceito
Declaração de Instituição e Infraestrutura	Carta_IAG.pdf	29/01/2016 09:12:24	ALINE SANTOS DAMASCENA	Aceito
Projeto Detalhado / Brochura Investigador	ProjetoAlineFev2016.pdf	29/01/2016 09:11:11	ALINE SANTOS DAMASCENA	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

SAO PAULO, 06 de Abril de 2016

Assinado por:
Maria Aparecida Azevedo Koike Folgueira
(Coordenador)

Referências

- Akaike H. A new look at the statistical model identification. *IEEE transactions on automatic control* 1974;19(6):716–723. [55](#)
- Al-Saadi J, Szykman J, Pierce RB, Kittaka C, Neil D, Chu DA, et al. Improving national air quality forecasts with satellite aerosol observations. *Bulletin of the American Meteorological Society* 2005;86(9):1249–1262. [2](#), [18](#)
- Albuquerque TTdA, West J, Andrade MdF, Ynoue RY, Andreão WL, dos Santos FS, et al. Analysis of PM 2.5 concentrations under pollutant emission control strategies in the metropolitan area of São Paulo, Brazil. *Environmental Science and Pollution Research* 2019;26(32):33216–33227. [11](#), [93](#)
- Albuquerque TTdA, Andrade MdF, Ynoue RY, Moreira DM, Andreão WL, dos Santos FS, et al. WRF-SMOKE-CMAQ modeling system for air quality evaluation in São Paulo megacity with a 2008 experimental campaign data. *Environmental Science and Pollution Research* 2018;25(36):36555–36569. [11](#), [93](#)
- Alden CB, Miller JB, Gatti LV, Gloor MM, Guan K, Michalak AM, et al. Regional atmospheric CO₂ inversion reveals seasonal and geographic differences in Amazon net biome exchange. *Global change biology* 2016;22(10):3427–3443. [30](#)
- de Almeida Castanho A, Prinn R, Martins V, Herold M, Ichoku C, Molina L. Analysis of Visible/SWIR surface reflectance ratios for aerosol retrievals from satellite in Mexico City urban area. *Atmospheric Chemistry and Physics* 2007;7:5467–5477. [87](#)
- de Almeida Castanho AD. Propriedades ópticas das partículas de aerossol e uma nova metodologia para a obtenção de espessura óptica via satélite sobre São Paulo. PhD thesis, São Paulo: Instituto de Física, Universidade de São Paulo; 2005. [18](#), [20](#), [31](#), [46](#), [88](#)
- de Almeida Castanho AD, Vanderlei Martins J, Artaxo P. MODIS aerosol optical depth retrievals with high spatial resolution over an urban area using the critical reflectance. *Journal of Geophysical Research: Atmospheres* 2008;113(D2). [50](#), [86](#), [87](#)
- Alonso MF, Longo KM, Freitas SR, da Fonseca RM, Marécal V, Pirre M, et al. An urban emissions inventory for South America and its application in numerical modeling of atmospheric chemical composition at local and regional scales. *Atmospheric Environment* 2010;44(39):5072–5083. [96](#)
- Alves E, Rodrigues C, Braga A, Miranda M, Nascimento Nd, Saldiva PN. Monitoramento da qualidade do ar no Brasil. São Paulo: Instituto Saúde e Sustentabilidade 2014;. [1](#), [15](#), [16](#), [17](#), [18](#)
- Amâncio CT, Nascimento LFC. Asma e poluentes ambientais: um estudo de séries temporais. *Revista da Associação Médica Brasileira* 2012;58(3):302–307. [5](#)

- Andersen ZJ, Bønnelykke K, Hvidberg M, Jensen SS, Ketzel M, Loft S, et al. Long-term exposure to air pollution and asthma hospitalisations in older adults: a cohort study. *Thorax* 2012;67(1):6–11. [25](#)
- Andrade MdF, Fornaro A, Mazzoli CR, Martins LD, Boian C, Oliveira MGL, et al. Ozone sounding in the Metropolitan Area of São Paulo, Brazil: Wet and dry season campaigns of 2006. *Atmospheric environment* 2012;61:627–640. [6](#), [25](#)
- Andrade MdF, Kumar P, de Freitas ED, Ynoue RY, Martins J, Martins LD, et al. Air quality in the megacity of São Paulo: Evolution over the last 30 years and future perspectives. *Atmospheric Environment* 2017;159:66–82. [15](#)
- Andrade MdF, de Miranda RM, Fornaro A, Kerr A, Oyama B, de Andre PA, et al. Vehicle emissions and PM 2.5 mass concentrations in six Brazilian cities. *Air Quality, Atmosphere & Health* 2012;5(1):79–88. [66](#)
- Andrade MdF, Ynoue RY, Freitas ED, Todesco E, Vara Vela A, Ibarra S, et al. Air quality forecasting system for Southeastern Brazil. *Frontiers in environmental Science* 2015;3:9. [11](#), [40](#), [93](#)
- Andrade Filho VSd, Artaxo Netto PE, Hacon SdS, Carmo CNd. Distribuição espacial de queimadas e mortalidade em idosos em região da Amazônia Brasileira, 2001–2012. *Ciência & Saúde Coletiva* 2017;22:245–253. [6](#)
- Andreão WL, Albuquerque TTdA. Fine particles as a public health indicator in Brazil: from monitoring to modeling. *Air Quality, Atmosphere & Health* 2020;p. 1–11. [6](#), [11](#), [15](#)
- Andreão WL, Pinto JA, Pedruzzi R, Kumar P, Albuquerque TTdA. Quantifying the impact of particle matter on mortality and hospitalizations in four Brazilian metropolitan areas. *Journal of Environmental Management* 2020;270:110840. [94](#)
- Ansmann A, Riebesell M, Weitkamp C. Measurement of atmospheric aerosol extinction profiles with a Raman lidar. *Optics letters* 1990;15(13):746–748. [49](#)
- Arbex MA, Martins LC, de Oliveira RC, Pereira LAA, Arbex FF, Cançado JED, et al. Air pollution from biomass burning and asthma hospital admissions in a sugar cane plantation area in Brazil. *Journal of Epidemiology & Community Health* 2007;61(5):395–400. [6](#)
- Bailey R, Barrie L, Halsall CJ, Fellin P, Muir D. Atmospheric organochlorine pesticides in the western Canadian Arctic: Evidence of transpacific transport. *Journal of Geophysical Research: Atmospheres* 2000;105(D9):11805–11811. [7](#)
- Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 2015;67(1):1–48. [46](#), [55](#), [96](#)
- Bell ML, Dominici F, Ebisu K, Zeger SL, Samet JM. Spatial and temporal variation in PM_{2.5} chemical composition in the United States for health effects studies. *Environmental health perspectives* 2007;115(7):989–995. [65](#), [66](#)
- Beloconi A, Chrysoulakis N, Lyapustin A, Utzinger J, Vounatsou P. Bayesian geostatistical modelling of PM₁₀ and PM_{2.5} surface level concentrations in Europe using high-resolution satellite-derived products. *Environment international* 2018;121:57–70. [95](#)

- Boyounk N, Léon JF, Delbarre H, Podvin T, Deroo C. Impact of the mixing boundary layer on the relationship between PM_{2.5} and aerosol optical thickness. *Atmospheric Environment* 2010;44(2):271–277. 49, 50
- Braga AL, Saldiva PH, Pereira LA, Menezes JJ, Conceição GM, Lin CA, et al. Health effects of air pollution exposure on children and adolescents in São Paulo, Brazil. *Pediatric pulmonology* 2001;31(2):106–113. 2
- Brasseur GP, Orlando JJ, Tyndall GS. *Atmospheric chemistry and global change*. Oxford university press; 1999. 34
- Brasseur GP, Prinn RG. *Atmospheric Chemistry in a Changing World: An Integration and Synthesis of a Decade of Tropospheric Chemistry Research; the International Global Atmospheric Chemistry Project of the International Geosphere-Biosphere Programme; with 22 Tables*. Springer Science & Business Media; 2003. 2, 11, 18
- Brunekreef B, Holgate ST. Air pollution and health. *The lancet* 2002;360(9341):1233–1242. 25
- Burke JM, Zufall MJ, OeZKAYNAK H. A population exposure model for particulate matter: case study results for PM_{2.5} in Philadelphia, PA. *Journal of Exposure Science & Environmental Epidemiology* 2001;11(6):470–489. 12
- Cançado JE, Saldiva PH, Pereira LA, Lara LB, Artaxo P, Martinelli LA, et al. The impact of sugar cane–burning emissions on the respiratory system of children and the elderly. *Environmental health perspectives* 2006;114(5):725–729. 6
- Carmo CNd, Hacon S, Longo KM, Freitas S, Ignotti E, Leon APd, et al. Associação entre material particulado de queimadas e doenças respiratórias na região sul da Amazônia brasileira. *Revista Panamericana de Salud Pública* 2010;27:10–16. 6
- Castanho AD, Artaxo P. Wintertime and summertime São Paulo aerosol source apportionment study. *Atmospheric Environment* 2001;35(29):4889–4902. 66
- Castanho ADA, Artaxo P, Martins JV, Hobbs PV, Remer L, Yamasoe M, et al. Chemical characterization of aerosols on the East Coast of the United States using aircraft and ground-based stations during the CLAMS experiment. *Journal of the atmospheric sciences* 2005;62(4):934–946. 66
- CETESB, Material particulado inalável fino (MP_{2,5}) e grosso (MP_{2,5} – 10) na atmosfera na Região Metropolitana de São Paulo (2000 – 2006); 2008. 29, 103
- CETESB, Classificação expedita da representatividade espacial das estações de monitoramento da qualidade do ar da CETESB no Estado de São Paulo - Terceira Etapa; 2016. 47, 48
- CETESB, Relatório Anual da Qualidade do Ar do Estado de São Paulo, 2015; 2016. 90
- CETESB, Relatório Anual da Qualidade do Ar do Estado de São Paulo, 2017; 2018. 14, 28
- CETESB, Relatório Anual da Qualidade do Ar do Estado de São Paulo, 2018; 2019. 3, 14, 25, 28, 49
- CETESB, Relatório de Emissões veiculares do Estado de São Paulo, 2018; 2019. 40, 41

- Chew BN, Campbell J, Hyer EJ, Salinas SV, Reid JS, Welton EJ, et al. Relationship between aerosol optical depth and particulate matter over Singapore: Effects of aerosol vertical distributions 2016;. 49
- Chouza F, Reitebuch O, Benedetti A, Weinzierl B. Saharan dust long-range transport across the Atlantic studied by an airborne Doppler wind lidar and the MACC model. *Atmospheric Chemistry & Physics* 2016;16(18). 7
- Christopherson RW. *Geossistemas-: Uma Introdução à Geografia Física*. Bookman Editora; 2012. 8
- Chu DA, Kaufman Y, Zibordi G, Chern J, Mao J, Li C, et al. Global monitoring of air pollution over land from the Earth Observing System-Terra Moderate Resolution Imaging Spectroradiometer (MODIS). *Journal of Geophysical Research: Atmospheres* 2003;108(D21). 2, 48
- Chu DA, Tsai TC, Chen JP, Chang SC, Jeng YJ, Chiang WL, et al. Interpreting aerosol lidar profiles to better estimate surface PM_{2.5} for columnar AOD measurements. *Atmospheric environment* 2013;79:172–187. 49
- Chu Y, Liu Y, Li X, Liu Z, Lu H, Lu Y, et al. A review on predicting ground PM_{2.5} concentration using satellite aerosol optical depth. *Atmosphere* 2016;7(10):129. 2, 18, 19, 21, 83
- Chudnovsky A, Lyapustin A, Wang Y, Tang C, Schwartz J, Koutrakis P. High resolution aerosol data from MODIS satellite for urban air quality studies. *Open Geosciences* 2014;6(1):17–26. 20, 89
- Chudnovsky AA, Koutrakis P, Kloog I, Melly S, Nordio F, Lyapustin A, et al. Fine particulate matter predictions using high resolution Aerosol Optical Depth (AOD) retrievals. *Atmospheric Environment* 2014;89:189–198. 12, 20, 59, 83, 91
- Clark M, Michael Clark: Mixed Models for Big Data; 2019. <https://m-clark.github.io/posts/2019-10-20-big-mixed-models/>. 96
- Conceição G, Miraglia S, Kishi HS, Saldiva P, Singer JM. Air pollution and child mortality: a time-series study in São Paulo, Brazil. *Environmental Health Perspectives* 2001;109(suppl 3):347–350. 5
- Daly C. Guidelines for assessing the suitability of spatial climate data sets. *International Journal of Climatology: A Journal of the Royal Meteorological Society* 2006;26(6):707–721. 93
- Damascena AS, Tanaka NI, Saldiva PHN. Multi-Angle Implementation of Atmospheric Correction (MAIAC) AOD product as a predictor of PM_{2.5} concentrations at São Paulo Metropolitan Region. In: *Proceedings of the Air Pollution Conference Brazil 4th CMAS South America Universidade Federal de Minas Gerais*; 2019. <https://cmasconference.com.br/wp-content/uploads/2019/10/Livro-Air-Pollution-2019-Final-V03-min.pdf>. 47
- Damascena AS, Tanaka NI, Saldiva PHN. Estimating daily PM_{2.5} concentrations at City of São Paulo, Brazil. In: *Proceedings of Abstracts 12th International Conference on Air Quality Science and Application University of Hertfordshire*; 2020. Doi: <https://doi.org/10.18745/pb.22217>. 57

- Day DE, Malm WC. Aerosol light scattering measurements as a function of relative humidity: a comparison between measurements made at three different sites. *Atmospheric Environment* 2001;35(30):5169–5176. 49
- DeVisser MH, Messina JP. Exploration of sensor comparability: a case study of composite MODIS Aqua and Terra data. *Remote sensing letters* 2013;4(6):599–608. 61
- Dockery DW, Pope CA, Xu X, Spengler JD, Ware JH, Fay ME, et al. An association between air pollution and mortality in six US cities. *New England journal of medicine* 1993;329(24):1753–1759. 4
- van der Does M, Brummer GJA, van Crimpen FC, Korte LF, Mahowald NM, Merkel U, et al. Tropical rains controlling deposition of Saharan dust across the North Atlantic Ocean. *Geophysical Research Letters* 2020;47(5):e2019GL086867. 7
- EEA. Air pollution: from emissions to exposure; 2016, <https://www.eea.europa.eu/media/infographics/air-pollution-from-emissions-to-exposure/view>, european Environmental Agency. 9
- Engel-Cox JA, Hoff RM, Rogers R, Dimmick F, Rush AC, Szykman JJ, et al. Integrating lidar and satellite optical depth with ambient monitoring for 3-dimensional particulate characterization. *Atmospheric Environment* 2006;40(40):8056–8067. 86, 88
- Engel-Cox JA, Holloman CH, Coutant BW, Hoff RM. Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. *Atmospheric environment* 2004;38(16):2495–2509. 85, 86
- Fajersztajn L, Veras M, Barrozo LV, Saldiva P. Air monitoring coverage in low-income countries: an observational study. *The Lancet* 2014;384:S14. 1, 18
- Fernandes MAdO, Andreão WL, Maciel FM, de Almeida Albuquerque TTdA. Avoiding hospital admissions for respiratory system diseases by complying to the final Brazilian air quality standard: an estimate for Brazilian southeast capitals. *Environmental Science and Pollution Research* 2020;p. 1–19. 6, 15
- Ferris BJ, Speizer F, Spengler J, Dockery D, Bishop Y, Wolfson M, et al. Effects of sulfur oxides and respirable particles on human health: methodology and demography of populations in study. *American Review of Respiratory Disease* 1979;120(4):767–779. 4
- Finlayson-Pitts BJ, Pitts Jr JN. *Chemistry of the upper and lower atmosphere: theory, experiments, and applications*. Elsevier; 1999. 3
- Freitas ED, Rozoff CM, Cotton WR, Dias PLS. Interactions of an urban heat island and sea-breeze circulations during winter over the metropolitan area of São Paulo, Brazil. *Boundary-layer meteorology* 2007;(1):43–65. 88
- Gallardo L, Escribano J, Dawidowski L, Rojas N, de Fátima Andrade M, Osses M. Evaluation of vehicle emission inventories for carbon monoxide and nitrogen oxides for Bogotá, Buenos Aires, Santiago, and São Paulo. *Atmospheric Environment* 2012;47:12–19. 96
- Gonçalves KdS, Castro HAd, Hacon SdS. As queimadas na região amazônica e o adoecimento respiratório. *Ciência & Saúde Coletiva* 2012;17:1523–1532. 6

- Gong W, Huang Y, Zhang T, Zhu Z, Ji Y, Xiang H. Impact and Suggestion of Column-to-Surface Vertical Correction Scheme on the Relationship between Satellite AOD and Ground-Level PM_{2.5} in China. *Remote Sensing* 2017;9(10):1038. [49](#)
- Gouveia N, Bremner S, Novaes H. Association between ambient air pollution and birth weight in São Paulo, Brazil. *Journal of Epidemiology & Community Health* 2004;58(1):11–17. [5](#)
- Gouveia N, Corrallo FP, Leon ACPd, Junger W, Freitas CUd. Poluição do ar e hospitalizações na maior metrópole brasileira. *Revista de Saúde Pública* 2017;51:117. [5](#)
- Gouveia N, Fletcher T. Respiratory diseases in children and outdoor air pollution in Sao Paulo, Brazil: a time series analysis. *Occupational and environmental medicine* 2000;57(7):477–483. [5](#)
- Grell GA, Peckham SE, Schmitz R, McKeen SA, Frost G, Skamarock WC, et al. Fully coupled “online” chemistry within the WRF model. *Atmospheric Environment* 2005;39(37):6957–6975. [93](#)
- Guinot B, Roger JC, Cachier H, Pucai W, Jianhui B, Tong Y. Impact of vertical atmospheric structure on Beijing aerosol distribution. *Atmospheric Environment* 2006;40(27):5167–5180. [48](#)
- Gupta P, Christopher SA, Wang J, Gehrig R, Lee Y, Kumar N. Satellite remote sensing of particulate matter and air quality assessment over global cities. *Atmospheric Environment* 2006;40(30):5880–5892. [40](#), [48](#)
- Gupta P, Remer LA, Levy RC, Mattoo S. Validation of MODIS 3 km land aerosol optical depth from NASA’s EOS Terra and Aqua missions. *Atmospheric Measurement Techniques* 2018;11(5):3145–3159. [20](#)
- Habermann M. Métodos de avaliação da exposição ambiental aos poluentes atmosféricos gerados por tráfego veicular. PhD thesis, São Paulo: Faculdade de Medicina, Universidade de São Paulo; 2012. [96](#)
- Habermann M, Gouveia NdC. Aplicação de regressão baseada no uso do solo para prever a concentração de material particulado inalável no município de São Paulo, Brasil. *Engenharia Sanitaria e Ambiental* 2012;17(2):155–162. [2](#), [40](#), [90](#), [96](#)
- Hadley O, Ramanathan V, Carmichael G, Tang Y, Corrigan C, Roberts G, et al. Trans-Pacific transport of black carbon and fine aerosols ($D < 2.5 \mu\text{m}$) into North America. *Journal of Geophysical Research: Atmospheres* 2007;112(D5). [7](#)
- Hänel G. The properties of atmospheric aerosol particles as functions of the relative humidity at thermodynamic equilibrium with the surrounding moist air. In: *Advances in geophysics*, vol. 19 Elsevier; 1976.p. 73–188. [49](#)
- Hartmann K, Krois J, Waske B. E-Learning Project SOGA: Statistics and Geospatial Data Analysis. Department of Earth Sciences, Freie Universitaet Berlin 2018;. [33](#)
- Henderson CR. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 1975;p. 423–447. [129](#)

- Hersey S, Garland RM, Crosbie E, Shingler T, Sorooshian A, Piketh S, et al. An overview of regional and local characteristics of aerosols in South Africa using satellite, ground, and modeling data. *Atmospheric chemistry and physics* (Print) 2015;15:4259. 85, 86
- Hoek G, Beelen R, De Hoogh K, Vienmeau D, Gulliver J, Fischer P, et al. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric environment* 2008;42(33):7561–7578. 11, 94
- Hoff RM, Christopher SA. Remote sensing of particulate pollution from space: have we reached the promised land? *Journal of the Air & Waste Management Association* 2009;59(6):645–675. 2, 18
- Holben BN, Eck TF, Slutsker I, Tanre D, Buis J, Setzer A, et al. AERONET—A federated instrument network and data archive for aerosol characterization. *Remote sensing of environment* 1998;66(1):1–16. 19
- Holben BN, Tanre D, Smirnov A, Eck T, Slutsker I, Abuhassan N, et al. An emerging ground-based aerosol climatology: Aerosol optical depth from AERONET. *Journal of Geophysical Research: Atmospheres* 2001;106(D11):12067–12097. 19
- de Hoogh K, Hritier H, Stafoggia M, Knzli N, Kloog I. Modelling daily PM_{2.5} concentrations at high spatio-temporal resolution across Switzerland. *Environmental Pollution* 2018;233:1147–1154. 20, 42, 59, 83, 84, 91, 92
- Hrub F, Fabianova E, Koppova K, Vandenberg JJ. Childhood respiratory symptoms, hospital admissions, and long-term exposure to airborne particulate matter. *Journal of Exposure Science and Environmental Epidemiology* 2001;11(1):33. 25
- Hsu N, Jeong MJ, Bettenhausen C, Sayer A, Hansell R, Seftor C, et al. Enhanced Deep Blue aerosol retrieval algorithm: The second generation. *Journal of Geophysical Research: Atmospheres* 2013;118(16):9296–9315. 88
- IEMA, 1 diagnstico das redes de monitoramento da qualidade do ar no Brasil; 2014. 15, 16, 17
- Ignotti E, Hacon SdS, Junger WL, Mouro D, Longo K, Freitas S, et al. Air pollution and hospital admissions for respiratory diseases in the subequatorial Amazon: a time series approach. *Cadernos de saude publica* 2010;26(4):747–761. 6
- Jin X, Fiore AM, Curci G, Lyapustin A, Civerolo K, Ku M, et al. Assessing uncertainties of a geophysical approach to estimate surface fine particulate matter distributions from satellite-observed aerosol optical depth. *Atmospheric Chemistry & Physics* 2019;19(1). 95
- Just AC, Wright RO, Schwartz J, Coull BA, Baccarelli AA, Tellez-Rojo MM, et al. Using high-resolution satellite aerosol optical depth to estimate daily PM_{2.5} geographical distribution in Mexico City. *Environmental science & technology* 2015;49(14):8576–8584. 20, 42, 55, 59, 83, 91, 92
- Kloog I, Chudnovsky AA, Just AC, Nordio F, Koutrakis P, Coull BA, et al. A new hybrid spatio-temporal model for estimating daily multi-year PM_{2.5} concentrations across northeastern USA using high resolution aerosol optical depth data. *Atmospheric Environment* 2014;95:581–590. 12

- Kloog I, Koutrakis P, Coull BA, Lee HJ, Schwartz J. Assessing temporally and spatially resolved PM_{2.5} exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmospheric environment* 2011;45(35):6267–6275. [12](#), [20](#), [59](#), [83](#), [91](#)
- Kloog I, Nordio F, Coull BA, Schwartz J. Incorporating local land use regression and satellite aerosol optical depth in a hybrid model of spatiotemporal PM_{2.5} exposures in the Mid-Atlantic states. *Environmental science & technology* 2012;46(21):11913–11921. [12](#), [20](#), [59](#), [83](#), [89](#), [91](#)
- Kloog I, Sorek-Hamer M, Lyapustin A, Coull B, Wang Y, Just AC, et al. Estimating daily PM_{2.5} and PM₁₀ across the complex geo-climate region of Israel using MAIAC satellite-based AOD data. *Atmospheric Environment* 2015;122:409–416. [20](#), [59](#), [83](#), [91](#)
- Knibbs LD, Van Donkelaar A, Martin RV, Bechle MJ, Brauer M, Cohen DD, et al. Satellite-based land-use regression for continental-scale long-term ambient PM_{2.5} exposure assessment in Australia. *Environmental science & technology* 2018;52(21):12445–12455. [95](#)
- Koelemeijer R, Homan C, Matthijsen J. Comparison of spatial and temporal variations of aerosol optical thickness and particulate matter over Europe. *Atmospheric Environment* 2006;40(27):5304–5315. [48](#), [49](#), [86](#)
- Kong L, Xin J, Zhang W, Wang Y. The empirical correlations between PM_{2.5}, PM₁₀ and AOD in the Beijing metropolitan region and the PM_{2.5}, PM₁₀ distributions retrieved by MODIS. *Environmental pollution* 2016;216:350–360. [49](#)
- Kutner MH, Nachtsheim CJ, Neter J, Li W, et al. *Applied linear statistical models*, vol. 5. McGraw-Hill Irwin New York; 2005. [37](#)
- Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 2017;82(13):1–26. [55](#)
- Laden F, Schwartz J, Speizer FE, Dockery DW. Reduction in fine particulate air pollution and mortality: extended follow-up of the Harvard Six Cities study. *American journal of respiratory and critical care medicine* 2006;173(6):667–672. [8](#)
- Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982;38(4):963–974. [12](#), [20](#), [22](#), [45](#), [89](#)
- LeBeau B. Misspecification of the random effect structure: Implications for the linear mixed model 2018;. [95](#)
- Lee H, Liu Y, Coull B, Schwartz J, Koutrakis P. A novel calibration approach of MODIS AOD data to predict PM_{2.5} concentrations. *Atmospheric Chemistry & Physics* 2011;11(15). [12](#), [20](#), [59](#), [83](#), [91](#)
- Lee HJ, Chatfield RB, Strawa AW. Enhancing the applicability of satellite remote sensing for PM_{2.5} estimation using MODIS deep blue AOD and land use regression in California, United States. *Environmental Science & Technology* 2016;50(12):6546–6555. [12](#), [20](#), [85](#), [86](#), [88](#), [91](#)
- Lee J, Kim J, Song C, Kim S, Chun Y, Sohn B, et al. Characteristics of aerosol types from AERONET sunphotometer measurements. *Atmospheric Environment* 2010;44(26):3110–3117. [51](#), [65](#), [66](#)

- Lepeule J, Laden F, Dockery D, Schwartz J. Chronic exposure to fine particles and mortality: an extended follow-up of the Harvard Six Cities study from 1974 to 2009. *Environmental health perspectives* 2012;120(7):965–970. 8
- Levy R, Mattoo S, Munchak L, Remer L, Sayer A, Patadia F, et al. The Collection 6 MODIS aerosol products over land and ocean. *Atmospheric Measurement Techniques* 2013;6(11):2989. 20, 88
- Levy R, Remer L, Kleidman R, Mattoo S, Ichoku C, Kahn R, et al. Global evaluation of the Collection 5 MODIS dark-target aerosol products over land. *Atmospheric Chemistry and Physics* 2010;10(21):10399–10420. 19, 61
- Levy RC, Remer LA, Mattoo S, Vermote EF, Kaufman YJ. Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance. *Journal of Geophysical Research: Atmospheres* 2007;112(D13). 19, 20, 50, 86, 87
- Li C, Mao J, Lau AK, Yuan Z, Wang M, Liu X. Application of MODIS satellite products to the air pollution research in Beijing. *Science in China Series D(Earth Sciences)* 2011;48:209–219. 49
- Li J, Carlson BE, Lacis AA. How well do satellite AOD observations represent the spatial and temporal variability of PM_{2.5} concentration for the United States? *Atmospheric Environment* 2015;102:260–273. 48, 86
- Lin C, Pereira L, Nishioka D, Conceição G, Braga A, Saldiva P. Air pollution and neonatal deaths in Sao Paulo, Brazil. *Brazilian Journal of Medical and Biological Research* 2004;37(5):765–770. 5
- Liu X, Cheng Y, Zhang Y, Jung J, Sugimoto N, Chang SY, et al. Influences of relative humidity and particle chemical composition on aerosol scattering properties during the 2006 PRD campaign. *Atmospheric Environment* 2008;42(7):1525–1536. 49
- Lopes AP, Nelson BW, Wu J, de Alencastro Graça PML, Tavares JV, Prohaska N, et al. Leaf flush drives dry season green-up of the Central Amazon. *Remote Sensing of Environment* 2016;182:90–98. 30
- Loría-Salazar SM, Holmes HA, Arnott WP, Barnard JC, Moosmüller H. Evaluation of MODIS columnar aerosol retrievals using AERONET in semi-arid Nevada and California, USA, during the summer of 2012. *Atmospheric Environment* 2016;144:345–360. 86
- Lyapustin A, Wang Y, Laszlo I, Kahn R, Korkin S, Remer L, et al. Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm. *Journal of Geophysical Research: Atmospheres* 2011;116(D3). 30, 87, 88
- Lyapustin A, Martonchik J, Wang Y, Laszlo I, Korkin S. Multiangle implementation of atmospheric correction (MAIAC): 1. Radiative transfer basis and look-up tables. *Journal of Geophysical Research: Atmospheres* 2011;116(D3). 50, 86
- Lyapustin A, Wang Y, Korkin S, Huang D. MODIS Collection 6 MAIAC algorithm. *Atmospheric Measurement Techniques* 2018;11(10). 20, 51, 61, 65, 87

- Macedo LR, Freitas E, Hoyos M, Rafee S. Relação entre temperatura e uso do solo: um estudo de caso para a RMSP. In: Anais do XIX Congresso Brasileiro de Meteorologia Congresso Brasileiro de Meteorologia; 2016. . 38
- Maeda EE, Ma X, Wagner FH, Kim H, Oki T, Eamus D, et al. Evapotranspiration seasonality across the Amazon Basin. *Earth System Dynamics* 2017;. 30
- Mariano G, Lopes F, Jorge M, Landulfo E. Assessment of biomass burnings activity with the synergy of sunphotometric and LIDAR measurements in São Paulo, Brazil. *Atmospheric Research* 2010;98(2-4):486–499. 7, 88
- Martins JV, Artaxo P, Kaufman YJ, Castanho AD, Remer LA. Spectral absorption properties of aerosol particles from 350–2500nm. *Geophysical research letters* 2009;36(13). 88
- Martins VS, Lyapustin A, de Carvalho LA, Barbosa CC, Novo EM. Validation of high-resolution MAIAC aerosol product over South America. *Journal of Geophysical Research: Atmospheres* 2017;122(14):7537–7559. 30, 46, 67, 87
- Martins Pereira G, Teinilä K, Custódio D, Gomes Santos A, Xian H, Hillamo R, et al. Particulate pollutants in the Brazilian city of São Paulo: 1-year investigation for the chemical composition and source apportionment. *Atmospheric Chemistry and Physics Discussions* 2017;17:11943–11969. 22, 40, 65, 66
- Medeiros A, Gouveia N. Relação entre baixo peso ao nascer e a poluição do ar no Município de São Paulo. *Revista de Saúde Pública* 2005;39:965–972. 5
- Meng J, Li C, Martin RV, van Donkelaar A, Hystad P, Brauer M. Estimated long-term (1981–2016) concentrations of ambient fine particulate matter across North America from chemical transport modeling, satellite remote sensing, and ground-based measurements. *Environmental science & technology* 2019;53(9):5071–5079. 95
- Millar G, Abel T, Allen J, Barn P, Noullett M, Spagnol J, et al. Evaluating human exposure to fine particulate matter part I: Measurements. *Geography Compass* 2010;4(4):281–302. 8, 9, 10
- Millar G, Abel T, Allen J, Barn P, Noullett M, Spagnol J, et al. Evaluating human exposure to fine particulate matter part II: Modeling. *Geography Compass* 2010;4(7):731–749. 9, 10, 11, 12, 94
- de Miranda RM, Lopes F, Do Rosário NÉ, Yamasoe MA, Landulfo E, de Fatima Andrade M. The relationship between aerosol particles chemical composition and optical properties to identify the biomass burning contribution to fine particles concentration: a case study for São Paulo city, Brazil. *Environmental monitoring and assessment* 2017;189(1):6. 7, 88
- Monn C. Exposure assessment of air pollutants: a review on spatial heterogeneity and indoor/outdoor/personal exposure to suspended particulate matter, nitrogen dioxide and ozone. *Atmospheric environment* 2001;35(1):1–32. 9, 10, 11, 94
- Moreira MA. Fundamentos do sensoriamento remoto e metodologias de aplicação. UFV; 2005. 38
- Morettin PA. Econometria financeira: um curso em séries temporais financeiras. Editora Blucher; 2017. 53

- Mota Lima FD, Pérez-Martínez PJ, de Fatima Andrade M, Kumar P, de Miranda RM. Characterization of particles emitted by pizzerias burning wood and briquettes: a case study at Sao Paulo, Brazil. *Environmental Science and Pollution Research* 2020;p. 1–14. 92
- Munchak L, Levy R, Mattoo S, Remer L, Holben B, Schafer J, et al. MODIS 3 km aerosol product: applications over land in an urban/suburban region 2013;. 20
- Nascimento AP, Santos JM, Mill JG, Souza JBd, Reis Júnior NC, Reisen VA. Associação entre concentração de partículas finas na atmosfera e doenças respiratórias agudas em crianças. *Revista de Saúde Pública* 2017;51:3. 5
- Nascimento LFC, Pereira LAA, Braga AL, Módolo MCC, Carvalho Jr JA. Efeitos da poluição atmosférica na saúde infantil em São José dos Campos, SP. *Revista de Saúde Pública* 2006;40:77–82. 5
- Natali L. Utilização dos produtos obtidos por sensoriamento remoto na caracterização da qualidade do ar na Região Metropolitana de São Paulo. PhD thesis, São Paulo: Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo; 2008. 2, 46, 47, 85
- NOAA-ESRL, SURFRAD Aerosol Optical Depth; 2020. <https://www.esrl.noaa.gov/gmd/grad/surfrad/aod/>, national Oceanic and Atmospheric Administration - Earth System Research Laboratories. 31
- Nobre JS, Singer JdM. Residual analysis for linear mixed models. *Biometrical Journal: Journal of Mathematical Methods in Biosciences* 2007;49(6):863–875. 95
- OECD. OECD Environmental Outlook to 2050: The Consequences of Inaction. Organization for Economic Co-operation and Development 2012;p. 207–273. 8
- Ott W, Thomas J, Mage D, Wallace L. Validation of the simulation of human activity and pollutant exposure (SHAPE) model using paired days from the Denver, CO, carbon monoxide field study. *Atmospheric Environment* (1967) 1988;22(10):2101–2113. 12
- Oudin A, Strömberg U, Jakobsson K, Stroh E, Lindgren AG, Norrving B, et al. Hospital admissions for ischemic stroke: does long-term exposure to air pollution interact with major risk factors? *Cerebrovascular Diseases* 2011;31(3):284–293. 25
- Pacheco MT, Parmigiani MMM, de Fatima Andrade M, Morawska L, Kumar P. A review of emissions and concentrations of particulate matter in the three major metropolitan areas of Brazil. *Journal of Transport & Health* 2017;4:53–72. 15
- Paciorek CJ, Liu Y, Moreno-Macias H, Kondragunta S. Spatiotemporal associations between GOES aerosol optical depth retrievals and ground-level PM2. 5. *Environmental science & technology* 2008;42(15):5800–5806. 85, 86
- Pedruzzi R, Baek BH, Henderson BH, Aravanis N, Pinto JA, Araujo IB, et al. Performance evaluation of a photochemical model using different boundary conditions over the urban and industrialized metropolitan area of Vitória, Brazil. *Environmental Science and Pollution Research* 2019;26(16):16125–16144. 93, 94

- Pereira Filho M, Pereira L, Arbex FF, Arbex M, Conceição G, Santos U, et al. Effect of air pollution on diabetes and cardiovascular diseases in São Paulo, Brazil. *Brazilian Journal of Medical and Biological Research* 2008;41(6):526–532. 5
- Pestana PRdS, Braga ALF, Ramos EMC, Oliveira AFd, Osadnik CR, Ferreira AD, et al. Effects of air pollution caused by sugarcane burning in Western São Paulo on the cardiovascular system. *Revista de Saúde Pública* 2017;51:13. 6
- Phalen RF, Phalen RN. Introduction to air pollution science: a public health perspective. Jones & Bartlett Publishers; 2011. 4
- Pinheiro J, Bates D. Mixed-effects models in S and S-PLUS. Springer Science & Business Media; 2000. 55
- Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. nlme: Linear and Nonlinear Mixed Effects Models; 2020, <https://CRAN.R-project.org/package=nlme>, r package version 3.1-140. 96
- Pope CA, Dockery DW, Schwartz J. Review of epidemiological evidence of health effects of particulate air pollution. *Inhalation toxicology* 1995;7(1):1–18. 25
- Prospero JM. Long-term measurements of the transport of African mineral dust to the southeastern United States: Implications for regional air quality. *Journal of Geophysical Research: Atmospheres* 1999;104(D13):15917–15927. 7
- Remer L, Mattoo S, Levy R, Munchak L. MODIS 3km aerosol product: algorithm and global perspective 2013;. 20
- Remer LA, Kaufman Y, Tanré D, Mattoo S, Chu D, Martins JV, et al. The MODIS aerosol algorithm, products, and validation. *Journal of the atmospheric sciences* 2005;62(4):947–973. 50
- Ribeiro AG, Downward GS, de Freitas CU, Neto FC, Cardoso MRA, de Oliveira MdRD, et al. Incidence and mortality for respiratory cancer and traffic-related air pollution in São Paulo, Brazil. *Environmental research* 2019;170:243–251. 2, 5
- Ribeiro H, Cardoso MRA. Air pollution and children's health in São Paulo (1986–1998). *Social science & medicine* 2003;57(11):2013–2022. 2
- Rodrigues PF. Avaliação da higroscopicidade de aerossóis urbanos pela técnica LIDAR Raman. PhD thesis, Universidade de São Paulo; 2014. 49, 50
- Romão R, Pereira LAA, Saldiva PHN, Pinheiro PM, Braga ALF, Martins LC. The relationship between low birth weight and exposure to inhalable particulate matter. *Cadernos de saúde pública* 2013;29:1101–1108. 5
- Saha S, Moorthi S, Wu X, Wang J, Nadiga S, Tripp P, et al. The NCEP climate forecast system version 2. *Journal of Climate* 2014;27(6):2185–2208. 36
- Saldiva P. Vida urbana e saúde: os desafios dos habitantes das metrópoles. Editora Contexto; 2018. 1
- Saldiva PH, King M, Delmonte V, Macchione M, Parada M, Daliberto M, et al. Respiratory alterations due to urban air pollution: an experimental study in rats. *Environmental Research* 1992;57(1):19–33. 5

- Saldiva PH, Lichtenfels A, Paiva P, Barone I, Martins M, Massad E, et al. Association between air pollution and mortality due to respiratory diseases in children in São Paulo, Brazil: a preliminary report. *Environmental research* 1994;65(2):218–225. [2](#), [5](#)
- Saldiva PH, Pope III CA, Schwartz J, Dockery DW, Lichtenfels AJ, Salge JM, et al. Air pollution and mortality in elderly people: a time-series study in Sao Paulo, Brazil. *Archives of Environmental Health: An International Journal* 1994;50(2):159–163. [2](#), [5](#)
- Saleska SR, Wu J, Guan K, Araujo AC, Huete A, Nobre AD, et al. Dry-season greening of Amazon forests. *Nature* 2016;531(7594):E4. [30](#)
- Sánchez MP, de Oliveira AP, Varona RP, Tito JV, Codato G, Ribeiro FND, et al. Rawinsonde-Based Analysis of the Urban Boundary Layer in the Metropolitan Region of São Paulo, Brazil. *Earth and Space Science* 2020;7(2). [34](#), [35](#)
- Sayer A, Hsu N, Bettenhausen C, Jeong MJ. Validation and uncertainty estimates for MODIS Collection 6 “Deep Blue” aerosol data. *Journal of Geophysical Research: Atmospheres* 2013;118(14):7864–7872. [20](#)
- Schaap M, Apituley A, Timmermans R, Koelemeijer R, Leeuw Gd. Exploring the relation between aerosol optical depth and PM 2.5 at Cabauw, the Netherlands. *Atmospheric Chemistry and Physics* 2009;9(3):909–925. [49](#)
- Seidel DJ, Ao CO, Li K. Estimating climatological planetary boundary layer heights from radiosonde observations: Comparison of methods and uncertainty analysis. *Journal of Geophysical Research: Atmospheres* 2010;115(D16). [35](#)
- Seinfeld JH, Pandis SN. *Atmospheric chemistry and physics: From air pollution to climate change*, John Willey & Sons; 2006. [3](#), [6](#), [7](#), [34](#)
- Silva FS, Cristale J, André PA, Saldiva PH, Marchi MR. PM2.5 and PM10: the influence of sugarcane burning on potential cancer risk. *Atmospheric Environment* 2010;44(39):5133–5138. [6](#)
- Singer JM, Nobre JS, Rocha FMM. Análise de dados longitudinais. versão preliminar; 2018. [22](#), [129](#)
- Singer JM, Rocha FM, Nobre JS. Graphical tools for detecting departures from linear mixed model assumptions and some remedial measures. *International Statistical Review* 2017;85(2):290–324. [95](#)
- Slater JF, Dibb JE. Relationships between surface and column aerosol radiative properties and air mass transport at a rural New England site. *Journal of Geophysical Research: Atmospheres* 2004;109(D1). [48](#)
- Sogacheva L, Popp T, Sayer AM, Dubovik O, Garay MJ, Heckel A, et al. Merging regional and global AOD records from 15 available satellite products 2019;. [19](#)
- Spix C, Anderson HR, Schwartz J, Vigotti MA, Letertre A, Vonk JM, et al. Short-term effects of air pollution on hospital admissions of respiratory diseases in Europe: a quantitative summary of APHEA study results. *Archives of Environmental Health: An International Journal* 1998;53(1):54–64. [25](#)

- Stafoggia M, Schwartz J, Badaloni C, Bellander T, Alessandrini E, Cattani G, et al. Estimation of daily PM₁₀ concentrations in Italy (2006–2012) using finely resolved satellite data, land use variables and meteorology. *Environment international* 2017;99:234–244. [20](#), [59](#), [83](#), [84](#), [91](#)
- Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, et al. Climate change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change 2013;1535. [7](#)
- Stull RB. An introduction to boundary layer meteorology, vol. 13. Springer Science & Business Media; 2012. [35](#)
- Takano APC, Justo LT, dos Santos NV, Marquezini MV, de André PA, da Rocha FMM, et al. Pleural anthracosis as an indicator of lifetime exposure to urban air pollution: An autopsy-based study in Sao Paulo. *Environmental research* 2019;173:23–32. [2](#)
- Tan Z. Air pollution and greenhouse gases: from basic concepts to engineering applications for air emission control. Springer; 2014. [3](#), [6](#), [7](#)
- Tao M, Wang J, Li R, Wang L, Wang L, Wang Z, et al. Performance of MODIS high-resolution MAIAC aerosol algorithm in China: Characterization and limitation. *Atmospheric environment* 2019;213:159–169. [86](#), [87](#)
- Thermo Scientific I. Instruction Manual for Thermo Scientific Model FH62C14 Beta Gauge Continuous Ambient Particulate Monitor. FH62C14 Instruction Manual 2010;. [49](#)
- Toth TD, Zhang J, Campbell JR, Hyer EJ, Reid JS, Shi Y, et al. Impact of data quality and surface-to-column representativeness on the PM_{2.5}/satellite AOD relationship for the contiguous United States. *Atmospheric Chemistry & Physics* 2014;14(12). [85](#), [86](#)
- Tsai TC, Jeng YJ, Chu DA, Chen JP, Chang SC. Analysis of the relationship between MODIS aerosol optical depth and particulate matter from 2006 to 2008. *Atmospheric Environment* 2011;45(27):4777–4788. [49](#)
- Tuan TS, Venâncio TS, Nascimento LFC. Air pollutants and hospitalization due to pneumonia among children. An ecological time series study. *Sao Paulo Medical Journal* 2015;133(5):408–413. [5](#)
- Vaida F, Blanchard S. Conditional Akaike information for mixed-effects models. *Biometrika* 2005;92(2):351–370. [56](#)
- Van Donkelaar A, Martin RV, Li C, Burnett RT. Regional estimates of chemical composition of fine particulate matter using a combined geoscience-statistical method with information from satellites, models, and monitors. *Environmental science & technology* 2019;53(5):2595–2611. [95](#)
- Vara-Vela A, de Fátima Andrade M, Zhang Y, Kumar P, Ynoue RY, Souto-Oliveira CE, et al. Modeling of Atmospheric Aerosol Properties in the São Paulo Metropolitan Area: Impact of Biomass Burning. *Journal of Geophysical Research: Atmospheres* 2018;123(17):9935–9956. [7](#), [11](#), [88](#), [93](#)
- Vardoulakis S, Fisher BE, Pericleous K, Gonzalez-Flesca N. Modelling air quality in street canyons: a review. *Atmospheric environment* 2003;37(2):155–182. [37](#)

- Vemado F, Pereira Filho AJ. Severe weather caused by heat island and sea breeze effects in the metropolitan area of São Paulo, Brazil. *Advances in Meteorology* 2016;2016. 88
- Veras MM, Caldini EG, Dolhnikoff M, Saldiva PHN. Air pollution and effects on reproductive-system functions globally with particular emphasis on the Brazilian population. *Journal of Toxicology and Environmental Health, Part B* 2010;13(1):1–15. 2
- Vormittag E, Rodrigues CG, Miranda Md, Cavalcante JA, Costa RRd, Camargo CA, et al. Avaliação do impacto da poluição atmosférica no Estado de São Paulo sob a visão da saúde. São Paulo: Instituto Saúde e Sustentabilidade 2013;. 5
- Vormittag E, Rodrigues CG, Miranda Md, Cavalcante JA, Costa RRd, Camargo CA, et al. Avaliação do impacto da poluição atmosférica no Estado de São Paulo sob a visão da saúde. São Paulo: Instituto Saúde e Sustentabilidade 2013;. 8
- Vormittag E, Saldiva P, Qualidade do Ar no Estado de São Paulo 2015 Sob a Visão da Saúde; 2017. 15
- Wagner FH, Hérault B, Rossi V, Hilker T, Maeda EE, Sanchez A, et al. Climate drivers of the Amazon forest greening. *PloS one* 2017;12(7):e0180932. 30
- Wallace JM, Hobbs PV. *Atmospheric science: an introductory survey*, vol. 92. Elsevier; 2006. 2, 3, 6, 18, 66
- Wang D, Morton D, Masek J, Wu A, Nagol J, Xiong X, et al. Impact of sensor degradation on the MODIS NDVI time series. *Remote Sensing of Environment* 2012;119:55–61. 61
- Wang J, Christopher SA. Intercomparison between satellite-derived aerosol optical thickness and PM_{2.5} mass: Implications for air quality studies. *Geophysical research letters* 2003;30(21). 2
- WHO. Air quality guidelines for Europe 2000;. 15
- WHO. Ambient air pollution: A global assessment of exposure and burden of disease 2016;. 1, 5, 8, 18, 30, 90
- WHO, et al. WHO guide to identifying the economic consequences of disease and injury 2009;. 7, 8
- Wilkening KE, Barrie LA, Engle M. Trans-Pacific air pollution. *Science* 2000;290(5489):65–67. 7
- WMO, WHO/IGAC Impacts of Megacities on Air Pollution and Climate; 2012. 40
- World Health Organization, Air Quality Guidelines: Global Update 2005: Particulate Matter, Ozone, Nitrogen Dioxide, and Sulfur Dioxide; 2006. 3, 4, 13, 83
- Xavier JCdM. Análise da disponibilidade da rede automática de monitoramento da qualidade do ar e seus reflexos no licenciamento ambiental realizado em São Paulo. PhD thesis, São José dos Campos: Instituto Tecnológico de Aeronáutica; 2011. 49
- Xiao Q, Wang Y, Chang HH, Meng X, Geng G, Lyapustin A, et al. Full-coverage high-resolution daily PM_{2.5} estimation using MAIAC AOD in the Yangtze River Delta of China. *Remote Sensing of Environment* 2017;199:437–446. 20, 59, 83, 91

- Yamasoe MA, do Rosário N, Barros K. Downward solar global irradiance at the surface in São Paulo city—The climatological effects of aerosol and clouds. *Journal of Geophysical Research: Atmospheres* 2017;122(1):391–404. [7](#), [88](#)
- Yu H, Chin M, Yuan T, Bian H, Remer LA, Prospero JM, et al. The fertilizing role of African dust in the Amazon rainforest: A first multiyear assessment based on data from Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations. *Geophysical Research Letters* 2015;42(6):1984–1991. [7](#)
- Zartarian VG, Ozkaynak H, Burke JM, Zufall MJ, Rigas ML, Furtaw Jr EJ. A modeling framework for estimating children’s residential exposure and dose to chlorpyrifos via dermal residue contact and nondietary ingestion. *Environmental Health Perspectives* 2000;108(6):505–514. [12](#)
- Zhang H, Hoff RM, Engel-Cox JA. The relation between Moderate Resolution Imaging Spectroradiometer (MODIS) aerosol optical depth and PM_{2.5} over the United States: a geographical comparison by US Environmental Protection Agency regions. *Journal of the Air & Waste Management Association* 2009;59(11):1358–1369. [85](#), [86](#)
- Zhang Q, Ma X, Tie X, Huang M, Zhao C. Vertical distributions of aerosols under different weather conditions: Analysis of in-situ aircraft measurements in Beijing, China. *Atmospheric Environment* 2009a;43(34):5526–5535. [48](#)
- Zhang Y, Song C, Band LE, Sun G, Li J. Reanalysis of global terrestrial vegetation trends from MODIS products: Browning or greening? *Remote Sensing of Environment* 2017;191:145–155. [61](#)
- Zhang Z, Wu W, Fan M, Wei J, Tan Y, Wang Q. Evaluation of MAIAC aerosol retrievals over China. *Atmospheric environment* 2019;202:8–16. [86](#), [87](#)
- Zheng C, Zhao C, Zhu Y, Wang Y, Shi X, Wu X, et al. Analysis of influential factors for the relationship between PM_{2.5} and AOD in Beijing. *Atmospheric Chemistry and Physics* 2017;17(21):13473–13489. [49](#)
- Zou B, Wilson JG, Zhan FB, Zeng Y. Air pollution exposure assessment methods utilized in epidemiological studies. *Journal of Environmental Monitoring* 2009;11(3):475–490. [8](#), [9](#), [10](#), [11](#), [12](#)