

UNIVERSIDADE DE SÃO PAULO
ESCOLA POLITÉCNICA DA UNIVERSIDADE DE SÃO PAULO
DEPARTAMENTO DE ENGENHARIA QUÍMICA

AMANDA CRISTALDO PIMENTEL

**Aplicação de análise estatística multivariada no estudo da distribuição espacial da
qualidade do ar na região metropolitana de São Paulo**

**Application of Multivariate Statistical Analysis in the Study of Spatial Distribution of
Air Quality in the Metropolitan Region of São Paulo**

São Paulo

2023

AMANDA CRISTALDO PIMENTEL

**Aplicação de análise estatística multivariada no estudo da distribuição espacial da
qualidade do ar na região metropolitana de São Paulo**

Versão Corrigida

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Química da Escola Politécnica da Universidade de São Paulo para a obtenção do título de Mestra em Ciências.

Área de Concentração: Engenharia Química

Orientador: Prof. Dr. Roberto Guardani


São Paulo


2023

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com anuência de seu orientador.

São Paulo, 2 de outubro de 2023

Assinatura do autor: 

Assinatura do orientador: 

Catálogo-na-publicação

Pimentel, Amanda Cristaldo

Aplicação de Análise Estatística Multivariada no Estudo da Distribuição Espacial da Qualidade do Ar na Região Metropolitana de São Paulo / A. C. Pimentel – versão corr. – São Paulo, 2023.

122 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo.

Departamento de Engenharia Química.

1.Qualidade do ar 2.Análise multivariada 3.Mapeamento I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia Química II.t.

Nome: PIMENTEL, Amanda Cristaldo

Título: Aplicação de análise estatística multivariada no estudo da distribuição espacial da qualidade do ar na região metropolitana de São Paulo

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Química da Escola Politécnica da Universidade de São Paulo para a obtenção do título de Mestra em Ciências.

Aprovado em: 15/09/2023

Banca Examinadora

Prof. Dr. Roberto Guardani

Instituição: Universidade de São Paulo (USP)

Julgamento: Aprovada

Profa. Dra. Maria de Fatima Andrade

Instituição: Universidade de São Paulo (USP)

Julgamento: Aprovada

Prof. Dra. Dirce Maria Pellegatti Franco

Instituição: CETESB

Julgamento: Aprovada

AGRADECIMENTOS

Ao Prof. Dr. Roberto Guardani, pela orientação deste estudo e por ser meu guia quando eu me perdia na imensidão dos dados. E principalmente pelas conversas motivadoras, que me permitiram seguir em frente e não desistir.

À minha família, que, mesmo depois de eu me mudar de volta para São Paulo, aceitou minhas visitas não tão frequentes, porque estava isolada em casa escrevendo a dissertação. Obrigada por me estimularem e serem por tantas vezes meu porto seguro.

Aos meus amigos dos grupos vocais Acappolli e DaCapo, que, nos meus momentos de estresse e desespero, compartilharam seu carinho, compreensão, e euforia contagiante.

Ao Thiago Varella, meu maior ídolo pesquisador, uma das pessoas mais brilhantes que já conheci, e que deu várias contribuições a este estudo, mesmo morando nos EUA.

Ao Thiago Kondo, meu melhor amigo, que me apoiou incondicionalmente nesta etapa da minha vida, assim como em todas as etapas dos 20 anos que se passaram desde que nos conhecemos.

À CETESB, pelo fornecimento dos dados utilizados no estudo e reuniões que ajudaram a nortear o trabalho.

À Universidade de São Paulo, por proporcionar um ambiente de constante estímulo intelectual e cultural, e reunir tantos pensadores brilhantes e com gana de fazer pesquisa, mesmo em situações tão adversas.

À Escola Politécnica, por continuar sendo a minha casa mesmo após a graduação, e por ter me tornado mais resiliente e uma exímia caçadora dos meus próprios sonhos.

“Em outras palavras, temos condições de assumir o controle do mecanismo do sucesso (...). Ainda assim, não fazemos isso. Por quê? Porque nos apegamos à ideia de que o sucesso é uma simples função do mérito individual e de que o mundo onde crescemos – e as regras que, como sociedade, optamos por criar – simplesmente não importa.” (Gladwell, 2008, p. 37)

RESUMO

CRISTALDO PIMENTEL, Amanda. **Aplicação de Análise Estatística Multivariada no Estudo da Distribuição Espacial da Qualidade do Ar na Região Metropolitana de São Paulo**. 122 f. Texto de qualificação de mestrado – Escola Politécnica, Universidade de São Paulo, São Paulo, 2023.

O monitoramento da qualidade do ar é importante para evitar prejuízos à vida da população, fauna e flora, além de representar gastos para o Estado. Para a gestão da qualidade do ar, se fazem necessárias ações de prevenção, combate e redução das emissões de poluentes e dos efeitos da degradação da atmosfera. Este estudo visa apresentar a aplicação de técnicas estatísticas de análise multivariada no tratamento de dados de qualidade do ar da Região Metropolitana de São Paulo (RMSP) utilizando-se bases de dados de estações de monitoramento da CETESB no período de cinco anos, 2016 a 2020, dos seguintes poluentes monitorados: Ozônio (O_3), partículas inaláveis e fumaça (MP_{10}) e partículas inaláveis finas ($MP_{2,5}$). O objetivo geral do estudo é identificar e quantificar correlações entre as estações monitoradas selecionadas, a fim de, por meio de medidas de similaridade, verificar a distribuição espacial dos principais poluentes atmosféricos na região, e identificar redundâncias entre as medições de emissões de poluentes. Assim, é possível propor a redução da quantidade de centros e/ou analisadores de medição, o que resultaria em uma redução de custos significativa em novos investimentos e manutenção. Foram utilizadas as seguintes técnicas estatísticas: matriz de correlação, análise de componentes principais (PCA), análise de fatores (AF) e análise de agrupamentos (clusters). Para aplicação dos algoritmos computacionais, escolheu-se a linguagem de programação Python. A conclusão do estudo demonstrou que os poluentes atmosféricos avaliados apresentam diferentes resultados devido às suas naturezas físicas e mecanismos de formação distintos. No caso do ozônio, é difícil concluir se existem medidores de fato redundantes, devido às altas correlações entre a maior parte das estações da RMSP. Para o MP_{10} , as estações Capão Redondo e Santo Amaro apresentaram maior potencial de terem um de seus analisadores removidos. Quanto ao $MP_{2,5}$, devido ao menor número de estações monitoradas na maior parte das análises, não é possível observar tendências de correlação entre as estações ao longo dos anos analisados.

Palavras-chave: Qualidade do ar. Análise multivariada. Mapeamento.

ABSTRACT

CRISTALDO PIMENTEL, Amanda. **Application of Multivariate Statistical Analysis in the Study of Spatial Distribution of Air Quality in the Metropolitan Region of São Paulo**. 122 p. Master's qualification text – Polytechnic School, University of São Paulo, São Paulo, 2023.

Monitoring air quality is important to prevent damage to the lives of the population, fauna and flora, in addition to representing expenses for the State. For the management of air quality, actions to prevent, combat and reduce pollutant emissions and the effects of atmospheric degradation are necessary. This study aims to present the application of statistical techniques of multivariate analysis in the treatment of air quality data in the Metropolitan Area of São Paulo (MASP) using databases from CETESB monitoring stations in the period of five years, 2016 to 2020, of the following monitored pollutants: Ozone (O_3), inhalable particles and smoke (PM_{10}) and fine inhalable particles ($PM_{2.5}$). The general objective of the study is to identify and quantify correlations between the selected monitoring stations, in order to, through similarity measures, verify the spatial distribution of the main atmospheric pollutants in the region, and identify redundancies between measurements of pollutant emissions. Thus, it is possible to propose a reduction in the number of monitoring stations and/or analisadors, which would result in a significant cost reduction in new investments and maintenance. The following statistical techniques were used: correlation matrix, principal component analysis (PCA), factor analysis (FA) and cluster analysis. For the application of computational algorithms, the Python programming language was chosen. The conclusion of the study demonstrated that the evaluated air pollutants present different results due to their different physical nature and formation mechanisms. In the case of ozone, it is difficult to conclude whether there are actually redundant analisadors, due to the high correlations between most stations in the MASP. For the PM_{10} , the Capão Redondo and Santo Amaro stations had the greatest potential to have one of their analisadors removed. As for $PM_{2.5}$, due to the smaller number of monitoring stations in most analyses, it is not possible to observe correlation trends between stations over the analyzed years.

Keywords: Air quality. Multivariate statistical analysis. Mapping.

LISTA DE ILUSTRAÇÕES

Figura 1 – Características, fontes e efeitos dos principais poluentes da atmosfera.....	14
Figura 2 - Principais efeitos à saúde para cada poluente e para cada classificação de qualidade do ar.....	16
Figura 3 - Rede automática com estações monitoras da RMSP (UGRHI 6) e variáveis medidas.....	17
Figura 4 - Localização das estações da Rede Automática na RMSP.....	18
Figura 5 - Evolução das emissões de poluentes veiculares na RMSP.....	22
Figura 6 - Mapa de representatividade do poluente MP _{2,5} na RMSP.....	26
Figura 7 - Mapa de representatividade do poluente MP ₁₀ na RMSP.....	27
Figura 8 - Mapa de representatividade do poluente O ₃ na RMSP.....	27
Figura 9 - Diferentes métodos de ligação para a Análise de Agrupamentos.....	34
Figura 10 - Gráfico de barras da quantidade de dados inválidos para cada estação de monitoramento da base de dados do ano 2020, poluente O ₃	40
Figura 11 - Mapa de calor (<i>heatmap</i>) da matriz de correlação da base de dados de 2016 a 2020 do poluente O ₃	43
Figura 12 - <i>Scree plot</i> da análise de componentes principais da base de dados de 2016 a 2020 do poluente O ₃	44
Figura 13 - <i>Scree plot</i> da análise de fatores da base de dados de 2016 a 2020 do poluente O ₃	46
Figura 14 - Dendrograma da análise de agrupamentos da base de dados de 2016 a 2020 do poluente O ₃	50
Figura 15 - Concentração média anual de O ₃ , em µg/m ³ , de 2016.....	52
Figura 16 - Concentração média anual de O ₃ , em µg/m ³ , de 2017.....	52
Figura 17 - Concentração média anual de O ₃ , em µg/m ³ , de 2018.....	53
Figura 18 - Concentração média anual de O ₃ , em µg/m ³ , de 2019.....	53
Figura 19 - Concentração média anual de O ₃ , em µg/m ³ , de 2020.....	54
Figura 20 - Concentração média de O ₃ , em µg/m ³ , no período quente e úmido em 2016.....	54
Figura 21 - Concentração média de O ₃ , em µg/m ³ , no período quente e úmido em 2017.....	55
Figura 22 - Concentração média de O ₃ , em µg/m ³ , no período quente e úmido em 2018.....	55
Figura 23 - Concentração média de O ₃ , em µg/m ³ , no período quente e úmido em 2019.....	56
Figura 24 - Concentração média de O ₃ , em µg/m ³ , no período quente e úmido em 2020.....	56

Figura 25 - Concentração média de O ₃ , em µg/m ³ , no período frio e seco em 2016.	57
Figura 26 - Concentração média de O ₃ , em µg/m ³ , no período frio e seco em 2017.....	57
Figura 27 - Concentração média de O ₃ , em µg/m ³ , no período frio e seco em 2018.....	58
Figura 28 - Concentração média de O ₃ , em µg/m ³ , no período frio e seco em 2019.....	58
Figura 29 - Concentração média de O ₃ , em µg/m ³ , no período frio e seco em 2020.....	59
Figura 30 - Número de períodos com alto grau de similaridade para cada par de estações de monitoramento de O ₃ , obtido na Análise de <i>Clusters</i>	68
Figura 31 - Concentração média anual de MP ₁₀ , em µg/m ³ , em 2016.	69
Figura 32 - Concentração média anual de MP ₁₀ , em µg/m ³ , em 2017.	70
Figura 33 - Concentração média anual de MP ₁₀ , em µg/m ³ , em 2018.	70
Figura 34 - Concentração média anual de MP ₁₀ , em µg/m ³ , em 2019.	71
Figura 35 - Concentração média anual de MP ₁₀ , em µg/m ³ , em 2020.	71
Figura 36 - Concentração média de MP ₁₀ , em µg/m ³ , no período quente e úmido em 2016. ..	72
Figura 37 - Concentração média de MP ₁₀ , em µg/m ³ , no período quente e úmido em 2017. ..	72
Figura 38 - Concentração média de MP ₁₀ , em µg/m ³ , no período quente e úmido em 2018. ..	73
Figura 39 - Concentração média de MP ₁₀ , em µg/m ³ , no período quente e úmido em 2019. ..	73
Figura 40 - Concentração média de MP ₁₀ , em µg/m ³ , no período quente e úmido em 2020. ..	74
Figura 41 - Concentração média de MP ₁₀ , em µg/m ³ , no período frio e seco em 2016.	74
Figura 42 - Concentração média de MP ₁₀ , em µg/m ³ , no período frio e seco em 2017.	75
Figura 43 - Concentração média de MP ₁₀ , em µg/m ³ , no período frio e seco em 2018.	75
Figura 44 - Concentração média de MP ₁₀ , em µg/m ³ , no período frio e seco em 2019.	76
Figura 45 - Concentração média de MP ₁₀ , em µg/m ³ , no período frio e seco em 2020.	76
Figura 46 - Número de períodos para cada par de estações de monitoramento de MP ₁₀ com alto grau de similaridade na Análise de <i>Clusters</i>	83
Figura 47 - Concentração média anual de MP _{2.5} , em µg/m ³ , em 2016.....	84
Figura 48 - Concentração média anual de MP _{2.5} , em µg/m ³ , em 2017.....	84
Figura 49 - Concentração média anual de MP _{2.5} , em µg/m ³ , em 2018.....	85
Figura 50 - Concentração média anual de MP _{2.5} , em µg/m ³ , em 2019.....	85
Figura 51 - Concentração média anual de MP _{2.5} , em µg/m ³ , em 2020.....	86
Figura 52 - Concentração média de MP _{2.5} , em µg/m ³ , no período quente e úmido em 2016...86	
Figura 53 - Concentração média de MP _{2.5} , em µg/m ³ , no período quente e úmido em 2017...87	
Figura 54 - Concentração média de MP _{2.5} , em µg/m ³ , no período quente e úmido em 2018...87	
Figura 55 - Concentração média de MP _{2.5} , em µg/m ³ , no período quente e úmido em 2019...88	

Figura 56 - Concentração média de MP _{2.5} , em µg/m ³ , no período quente e úmido em 2020...	88
Figura 57 - Concentração média de MP _{2.5} , em µg/m ³ , no período frio e seco e, 2016.....	89
Figura 58 - Concentração média de MP _{2.5} , em µg/m ³ , no período frio e seco e, 2017.....	89
Figura 59 - Concentração média de MP _{2.5} , em µg/m ³ , no período frio e seco e, 2018.....	90
Figura 60 - Concentração média de MP _{2.5} , em µg/m ³ , no período frio e seco e, 2019.....	90
Figura 61 - Concentração média de MP _{2.5} , em µg/m ³ , no período frio e seco e, 2020.....	91
Figura 62 - Número de períodos para cada par de estações de monitoramento de MP _{2.5} com alto grau de similaridade na Análise de <i>Clusters</i>	98
Figura 63 - Dados de temperatura (°C) no período de 2016 a 2020.....	103
Figura 64 - Dados de umidade relativa do ar (%) no período de 2016 a 2020.....	103
Figura 65 - Representação gráfica da matriz de correlação do poluente O ₃ no período de 2016 a 2020.	107
Figura 66 - Representação gráfica da matriz de correlação do poluente MP ₁₀ no período de 2016 a 2020.	107
Figura 67 - Representação gráfica da matriz de correlação do poluente MP _{2.5} no período de 2016 a 2020.	107
Figura 68 - Scree plot da PCA do poluente O ₃ no período de 2016 a 2020.....	108
Figura 69 - Scree plot do PCA do poluente MP ₁₀ no período de 2016 a 2020.....	109
Figura 70 - Scree plot do PCA do poluente MP _{2.5} no período de 2016 a 2020.....	111
Figura 71 - Scree plot da análise de fatores do poluente O ₃ no período de 2016 a 2020.....	112
Figura 72 - Scree plot da análise de fatores do poluente MP ₁₀ no período de 2016 a 2020...	113
Figura 73 - Scree plot da análise de fatores do poluente MP _{2.5} no período de 2016 a 2020. .	114
Figura 74 - Dendrograma da análise de clusters do poluente O ₃ no período de 2016 a 2020.	115
Figura 75 - Dendrograma da análise de clusters do poluente MP ₁₀ no período de 2016 a 2020.	115
Figura 76 - Dendrograma da análise de clusters do poluente MP _{2.5} no período de 2016 a 2020.	116

LISTA DE TABELAS

Tabela 1 - Definição de escalas de representatividade espacial.	24
Tabela 2 - Tipos de dispersão para análises de agrupamentos.	33
Tabela 3 - Variáveis contidas na base de dados original da CETESB.	36
Tabela 4 - Significado e tipo de cálculo das variáveis contidas na base de dados da CETESB (CETESB, 2021).	36
Tabela 5 - Períodos de análise das bases de dados dos poluentes O ₃ , MP ₁₀ e MP _{2,5}	37
Tabela 6 - Legenda dos códigos utilizados nas bases de dados para as estações automáticas fixas da RMSB para os poluentes MP ₁₀ , MP _{2,5} e O ₃	39
Tabela 7 - Estações eliminadas de cada base de dados devido à alta quantidade de dados inválidos. Por período e por poluente.	40
Tabela 8 - Resumo dos principais resultados das técnicas de análise estatística multivariada para o poluente O ₃	60
Tabela 9 - Resumo dos principais resultados das técnicas de análise estatística multivariada para o poluente MP ₁₀	77
Tabela 10 - Resumo dos principais resultados das técnicas de análise estatística multivariada para o poluente MP _{2,5}	91
Tabela 11 – Matriz de correlação do poluente O ₃ no período de 2016 a 2020.	104
Tabela 12 - Matriz de correlação do poluente MP ₁₀ no período de 2016 a 2020.	105
Tabela 13 - Matriz de correlação do poluente MP _{2,5} no período de 2016 a 2020.	106
Tabela 14 - Tabela resumo da PCA do poluente O ₃ no período de 2016 a 2020.	108
Tabela 15 – Pesos das variáveis sobre o componente 1, em ordem decrescente, do poluente O ₃ no período de 2016 a 2020.	109
Tabela 16 - Tabela resumo da PCA do poluente MP ₁₀ no período de 2016 a 2020.	110
Tabela 17 - Pesos das variáveis sobre o componente 1, em ordem decrescente, do poluente MP ₁₀ no período de 2016 a 2020.	110
Tabela 18 - Tabela resumo da PCA do poluente MP _{2,5} no período de 2016 a 2020.	111
Tabela 19 - Pesos das variáveis sobre o componente 1, em ordem decrescente, do poluente MP _{2,5} no período de 2016 a 2020.	111
Tabela 20 - Cargas fatoriais do poluente O ₃ no período de 2016 a 2020 em ordem decrescente.	112

Tabela 21 - Cargas fatorias do poluente MP ₁₀ no período de 2016 a 2020 em ordem decrescente.....	113
Tabela 22 - Cargas fatorias do poluente MP _{2,5} no período de 2016 a 2020 em ordem decrescente.....	114
Tabela 23 - Cargas fatorias do poluente MP ₁₀ , período quente e úmido do ano de 2016.	117
Tabela 24 - Cargas fatorias do poluente MP ₁₀ , período quente e úmido do ano de 2017.	117
Tabela 25 - Cargas fatorias do poluente MP ₁₀ , período quente e úmido do ano de 2018.	117
Tabela 26 - Cargas fatorias do poluente MP ₁₀ , período quente e úmido do ano de 2019.	118
Tabela 27 - Cargas fatorias do poluente MP ₁₀ , período quente e úmido do ano de 2020.	118
Tabela 28 - Cargas fatorias do poluente MP _{2,5} , período quente e úmido do ano de 2018. ...	119
Tabela 29 - Cargas fatorias do poluente MP _{2,5} , período quente e úmido do ano de 2019. ...	119

SUMÁRIO

1. INTRODUÇÃO	13
1.1. MOTIVAÇÃO	13
1.2. OBJETIVOS	18
1.3. ESTRUTURA DA DISSERTAÇÃO	19
2. REVISÃO BIBLIOGRÁFICA	20
2.1. QUALIDADE DO AR NA REGIÃO METROPOLITANA DE SÃO PAULO	20
2.2. DESCRIÇÃO DA REDE DE MONITORAMENTO DA CETESB NA RMSP	23
2.3. MÉTODOS DE ESTATÍSTICA MULTIVARIADA	28
2.3.1. <i>Matriz de Correlação</i>	29
2.3.2. <i>Análise de Componentes Principais</i>	29
2.3.3. <i>Análise de Fatores</i>	31
2.3.4. <i>Análise de Agrupamentos ou Clusters</i>	32
3. METODOLOGIA	35
3.1. MATRIZ DE CORRELAÇÃO	42
3.2. ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)	43
3.3. ANÁLISE DE FATORES (AF)	44
3.4. ANÁLISE DE AGRUPAMENTOS OU <i>CLUSTERS</i>	47
4. RESULTADOS E DISCUSSÃO	51
4.1. O ₃	51
4.2. MP ₁₀	69
4.3. MP _{2,5}	83
5. CONCLUSÕES	99
6. REFERÊNCIAS BIBLIOGRÁFICAS	101
APÊNDICE A – SELEÇÃO DE PERÍODOS PELA TEMPERATURA E UMIDADE RELATIVA	103
APÊNDICE B – EXEMPLOS DE MATRIZES DE CORRELAÇÃO	104

APÊNDICE C – EXEMPLOS DE ANÁLISES DE COMPONENTES PRINCIPAIS (PCA)	108
APÊNDICE D – EXEMPLOS DE ANÁLISES DE FATORES	112
APÊNDICE E – EXEMPLOS DE ANÁLISES DE AGRUPAMENTOS (<i>CLUSTERS</i>)	115
APÊNDICE F – CARGAS FATORIAIS DAS ANÁLISES DE FATORES DO POLUENTE MP₁₀, PERÍODOS QUENTES E ÚMIDOS COM MAIS DE 1 FATOR	117
APÊNDICE G – CARGAS FATORIAIS DAS ANÁLISES DE FATORES DO POLUENTE MP_{2,5}, PERÍODOS QUENTES E ÚMIDOS COM MAIS DE 1 FATOR	119

1. INTRODUÇÃO

1.1. Motivação

Assim como outras megacidades, a Região Metropolitana de São Paulo (RMSP) sofre de problemas relativos à qualidade do ar, sendo o elevado número de automóveis sua maior influência. Os efeitos nocivos das emissões veiculares derivadas da queima de combustível fóssil estão entre as principais razões pelas quais desenvolvem-se estudos sobre os fatores que influem na qualidade do ar, e criam-se políticas com o objetivo de reduzir as emissões. O programa mais eficaz para a redução de emissões veiculares foi a implementação, em 1986, em nível nacional, do PROCONVE, Programa de Controle da Poluição do Ar por Veículos Automotores, que desde então aplicou uma política de estabelecer limites para essas emissões (ANDRADE *et al.*, 2017).

O nível de poluição atmosférica é determinado pela concentração de substâncias poluentes presentes no ar. Conforme a RESOLUÇÃO CONAMA Nº 491 DE 19/11/2018 (CONAMA, 2018), poluente atmosférico é “qualquer forma de matéria em quantidade, concentração, tempo ou outras características, que tornem ou possam tornar o ar impróprio ou nocivo à saúde, inconveniente ao bem-estar público, danoso aos materiais, à fauna e flora ou prejudicial à segurança, ao uso e gozo da propriedade ou às atividades normais da comunidade”.

Os poluentes mais utilizados como indicadores da qualidade do ar são monóxido de carbono (CO), dióxido de enxofre (SO₂), dióxido de nitrogênio (NO₂), material particulado e ozônio (O₃). Tanto a maior frequência desses poluentes na atmosfera quanto os efeitos nocivos à saúde e o meio ambiente são motivos para a escolha desses parâmetros.

Os poluentes podem ser classificados, de acordo com a sua origem, como:

- Primários: emitidos diretamente pelas fontes de emissão;
- Secundários: formados na atmosfera através de reações químicas entre poluentes e/ou constituintes naturais na atmosfera.

Os poluentes, assim como são provenientes de diferentes fontes, também apresentam diferentes efeitos no meio ambiente. A Figura 1 apresenta uma tabela extraída do relatório anual de qualidade do ar da Companhia Ambiental do Estado de São Paulo (CETESB) do ano de 2020 com as principais características, fontes e efeitos de cada poluente no meio ambiente.

Poluente	Características	Fontes Principais	Efeitos Gerais ao Meio Ambiente
Partículas Inaláveis Finas (MP _{2,5})	Partículas de material sólido ou líquido suspensas no ar, na forma de poeira, neblina, aerossol, fumaça, fuligem etc., que podem permanecer no ar e percorrer longas distâncias. Faixa de tamanho ≤ 2,5 micra.	Processos de combustão (industrial, veículos automotores), aerossol secundário (formado na atmosfera) como sulfato e nitrato, entre outros.	Danos à vegetação, deterioração da visibilidade e contaminação do solo e da água.
Partículas Inaláveis (MP ₁₀) e Fumaça	Partículas de material sólido ou líquido que ficam suspensas no ar, na forma de poeira, neblina, aerossol, fumaça, fuligem etc. Faixa de tamanho ≤ 10 micra.	Processos de combustão (indústria e veículos automotores), poeira ressuspensa, aerossol secundário (formado na atmosfera).	Danos à vegetação, deterioração da visibilidade e contaminação do solo e da água.
Partículas Totais em Suspensão (PTS)	Partículas de material sólido ou líquido que ficam suspensas no ar, na forma de poeira, neblina, aerossol, fumaça, fuligem etc. Faixa de tamanho ≤ 50 micra.	Processos industriais, veículos motorizados (exaustão), poeira de rua ressuspensa, queima de biomassa. Fontes naturais: pólen, aerossol marinho e solo.	Danos à vegetação, deterioração da visibilidade e contaminação do solo e da água.
Dióxido de Enxofre (SO ₂)	Gás incolor, com forte odor, semelhante ao gás produzido na queima de palitos de fósforos. Pode ser oxidado a SO ₃ , que na presença de vapor de água, passa rapidamente a H ₂ SO ₄ . É um importante precursor dos sulfatos, um dos principais componentes das partículas inaláveis.	Processos que utilizam queima de óleo combustível, refinarias de petróleo, veículos a diesel, produção de polpa de celulose e papel, fertilizantes.	Pode levar à formação de chuva ácida, causar corrosão aos materiais e danos à vegetação: folhas e colheitas.
Dióxido de Nitrogênio (NO ₂)	Gás marrom avermelhado, com odor forte e muito irritante. Pode levar à formação de ácido nítrico, nitratos (os quais contribuem para o aumento das partículas inaláveis na atmosfera) e compostos orgânicos tóxicos.	Processos de combustão envolvendo veículos automotores, processos industriais, usinas térmicas, incinerações.	Pode levar à formação de chuva ácida, danos à vegetação e à colheita.
Monóxido de Carbono (CO)	Gás incolor, inodoro e insípido.	Combustão incompleta em veículos automotores.	
Ozônio (O ₃)	Gás incolor, inodoro nas concentrações ambientais e o principal componente da névoa fotoquímica.	Não é emitido diretamente para a atmosfera. É produzido fotoquimicamente pela radiação solar sobre os óxidos de nitrogênio e compostos orgânicos voláteis.	Danos às colheitas, à vegetação natural, plantações agrícolas; às plantas ornamentais.

Figura 1 – Características, fontes e efeitos dos principais poluentes da atmosfera.

Fonte: (CETESB, 2021)

De acordo com dados da CETESB, o ozônio apresenta ao longo do ano uma distribuição de episódios totalmente distinta dos poluentes primários, uma vez que é um poluente secundário, ou seja, é formado na atmosfera através de reações fotoquímicas, dependentes da incidência de luz solar e presença de precursores. De maneira geral, no Estado de São Paulo, as concentrações mais altas de ozônio são observadas no período de primavera e verão (entre setembro e março), principalmente em dias sem ou com pouca nebulosidade, o que permite a incidência de radiação solar (CETESB, 2021).

Na RMSP há grande emissão de precursores do ozônio, principalmente de origem veicular, como compostos orgânicos voláteis (COV) e óxidos de nitrogênio (NO_x), apresentando um alto potencial de formação desse poluente. Porém, a variação das concentrações de ozônio está relacionada, principalmente, às mudanças das condições meteorológicas, pois não há significativas variações quantitativas nas emissões de seus precursores de ano para ano. Além disso, não é possível observar uma tendência na concentração de ozônio ao longo dos anos, devido às interações químicas e meteorológicas

complexas nas reações atmosféricas de formação e transporte deste poluente (ANDRADE *et al.*, 2012).

No caso da RMSP, o quadro de altas concentrações de ozônio conduz à necessidade de maior controle dos precursores da formação deste poluente por processos fotoquímicos. Tais processos geram mais substâncias agressivas além do ozônio, denominadas genericamente de oxidantes fotoquímicos, e contribuem para a formação de aerossóis secundários, que em função de seu tamanho reduzido podem penetrar profundamente no sistema respiratório, afetando a saúde da população (CETESB, 2021).

A grande complexidade dos processos de formação de poluentes secundários na atmosfera tem motivado, desde o final da década de 1990, estudos voltados à previsão da qualidade do ar na RMSP com base em modelos estatísticos multivariados, em estudos conjuntos entre a USP e a CETESB.

Quanto às partículas inaláveis, MP_{10} e $MP_{2,5}$, trata-se principalmente de poluentes primários que são emitidos diretamente por fontes de emissão, como veículos automotores. Segundo Andrade *et al.* (ANDRADE *et al.*, 2012), através da aplicação de modelos receptores para concentrações elementares de partículas finas em seis capitais do Brasil (São Paulo, Rio de Janeiro, Curitiba, Belo Horizonte, Recife e Porto Alegre), constataram que as emissões veiculares são a principal fonte de $MP_{2,5}$ nas maiores cidades do país. Como indicado na Figura 1, devido à forma como são classificados de acordo com o tamanho das partículas, o $MP_{2,5}$ é compreendido pelo MP_{10} , e corresponde a cerca de 60% deste, segundo medidas realizadas pela CETESB na RMSP desde 1987.

Em decorrência de ações e programas de controles de emissões, verifica-se uma melhora nos níveis de concentração de MP_{10} na RMSP quando se comparam as concentrações atuais com as do início da década de 2000. Observa-se, também, que as concentrações médias deste poluente tendem à estabilidade, o que se deve às emissões cada vez menores dos veículos novos combinada ao aumento da frota (CETESB, 2021).

A Figura 2 ilustra para as diferentes classificações de qualidade do ar, que é baseada no cálculo de um índice, quais são os efeitos à saúde da população para cada poluente.

Qualidade	Índice	MP ₁₀ (µg/m ³) 24h	MP _{2,5} (µg/m ³) 24h	O ₃ (µg/m ³) 8h	CO (ppm) 8h	NO ₂ (µg/m ³) 1h	SO ₂ (µg/m ³) 24h
N1 - BOA	0 - 40	0 - 50	0 - 25	0 - 100	0 - 9	0 - 200	0 - 20
N2 - MODERADA	41 - 80	>50 - 100 Pessoas com doenças respiratórias podem apresentar sintomas, como tosse seca e cansaço.	>25 - 50 Pessoas com doenças respiratórias podem apresentar sintomas, como tosse seca e cansaço.	>100 - 130 Pessoas com doenças respiratórias podem apresentar sintomas, como tosse seca e cansaço.	>9 - 11 Pessoas com doenças cardíacas podem apresentar sintomas, como cansaço e dor no peito.	>200 - 240 Pessoas com doenças respiratórias podem apresentar sintomas, como tosse seca e cansaço.	>20 - 40 Pessoas com doenças respiratórias podem apresentar sintomas, como tosse seca e cansaço.
N3 - RUIM	81 - 120	>100 - 150 Pessoas com doenças respiratórias ou cardíacas, idosos e crianças têm os sintomas agravados. População em geral pode apresentar sintomas, como ardor nos olhos, nariz e garganta, tosse seca e cansaço.	>50 - 75 Pessoas com doenças respiratórias ou cardíacas, idosos e crianças têm os sintomas agravados. População em geral pode apresentar sintomas, como ardor nos olhos, nariz e garganta, tosse seca e cansaço.	>130 - 160 Pessoas com doenças respiratórias, como asma, e crianças têm os sintomas agravados. População em geral pode apresentar sintomas, como ardor nos olhos, nariz e garganta, tosse seca e cansaço.	>11 - 13 População em geral pode apresentar sintomas, como cansaço. Pessoas com doenças cardíacas têm os sintomas como cansaço e dor no peito agravados.	>240 - 320 População em geral pode apresentar sintomas, como ardor nos olhos, nariz e garganta, tosse seca e cansaço. Pessoas com doenças respiratórias e crianças têm os sintomas agravados.	>40 - 365 População em geral pode apresentar sintomas, como ardor nos olhos, nariz e garganta, tosse seca e cansaço. Pessoas com doenças respiratórias ou cardíacas, idosos e crianças têm os sintomas agravados.
N4 - MUITO RUIM	121-200	>150 - 250 Aumento dos sintomas em crianças e pessoas com doenças pulmonares e cardiovasculares. Aumento de sintomas respiratórios na população em geral.	>75 - 125 Aumento dos sintomas em crianças e pessoas com doenças pulmonares e cardiovasculares. Aumento de sintomas respiratórios na população em geral.	>160 - 200 Aumento dos sintomas respiratórios em crianças e pessoas com doenças pulmonares, como asma. Aumento de sintomas respiratórios na população em geral.	>13 - 15 Aumento de sintomas em pessoas cardíacas. Aumento de sintomas cardiovasculares na população em geral.	>320 - 1130 Aumento dos sintomas respiratórios em crianças e pessoas com doenças pulmonares, como asma. Aumento de sintomas respiratórios na população em geral.	>365 - 800 Aumento dos sintomas em crianças e pessoas com doenças pulmonares e cardiovasculares. Aumento de sintomas respiratórios na população em geral.
N5 - PÉSSIMA	>200	>250 Agravamento dos sintomas respiratórios. Agravamento de doenças pulmonares, como asma, e cardiovasculares, como infarto do miocárdio.	>125 Agravamento dos sintomas respiratórios. Agravamento de doenças pulmonares, como asma, e cardiovasculares, como infarto do miocárdio.	>200 Agravamento de sintomas respiratórios. Agravamento de doenças pulmonares, como asma, e doença pulmonar obstrutiva crônica.	>15 Agravamento das doenças cardiovasculares, como infarto do miocárdio e insuficiência cardíaca congestiva.	>1130 Agravamento de sintomas respiratórios. Agravamento de doenças pulmonares, como asma, e doença pulmonar obstrutiva crônica.	>800 Agravamento dos sintomas respiratórios. Agravamento de doenças pulmonares, como asma, e cardiovasculares, como infarto do miocárdio.

Figura 2 - Principais efeitos à saúde para cada poluente e para cada classificação de qualidade do ar.

Fonte: (CETESB, 2021).

A Figura 3 apresenta as estações monitoras da RMSP, bem como quais variáveis são medidas em cada uma. A tabela consiste em um recorte extraído do relatório anual de qualidade do ar da CETESB do ano de 2021.

Localização das Estações	PARÂMETROS																
	MP2,5	MP10	SOx	NO	NO2	NOx	CO	O3	BEN	TOL	ERT	UR	TEMP	VV	DV	P	RAD
Capão Redondo		X						X				X	X	X	X	X	X
Carapicuíba		X					X	X				X	X	X	X	X	X
Cerqueira César		X	X	X	X	X	X	X									
Cid. Universitária - USP - IPEN	X							X									
Congonhas	X	X	X	X	X	X	X	X									
Diadema		X						X									
Grajaú-Parelheiros	X	X					X	X			X	X					
Guarulhos - Paço Municipal	X	X						X			X	X	X	X	X	X	X
Guarulhos - Pimentas	X	X	X	X	X	X	X	X			X	X	X	X	X	X	X
Ibirapuera	X			X	X	X	X	X			X	X				X	X
Interlagos		X	X	X	X	X		X			X	X	X	X	X	X	X
Itaim Paulista	X	X		X	X	X		X					X	X			
Itaquera								X									
Marg. Tietê - Pte. dos Remédios	X	X	X	X	X	X	X	X			X	X	X	X	X	X	X
Mauá	X	X						X									
Mooca	X						X	X					X	X			
Nossa Senhora do Ó		X						X			X	X					
Osasco	X	X	X	X	X	X	X	X					X	X	X	X	
Parque D. Pedro II	X	X		X	X	X	X	X			X	X	X	X	X	X	X
Pico do Jaraguá	X			X	X	X		X			X	X	X	X			
Pinheiros	X	X		X	X	X	X	X	X	X	X	X	X	X	X		
Santana	X							X						X	X		
Santo Amaro		X		X	X	X	X	X						X	X		
S. André - Capuava		X	X	X	X	X	X	X	X	X				X	X		
S. Bernardo - Centro	X			X	X	X	X	X			X	X	X	X	X	X	X
S. Bernardo - Pauliceia		X															
S. Caetano	X	X	X	X	X	X	X	X			X	X	X	X	X	X	X
Taboão da Serra		X						X			X	X					
TOTAL MONITORES FIXOS UGRNI 6	17	21	8	15	15	15	15	22	2	2	1	15	15	17	17	10	8

MP _{2,5}	Partículas inaláveis finas	CO	Monóxido de carbono	UR	Umidade relativa do ar
MP ₁₀	Partículas inaláveis	O ₃	Ozônio	TEMP	Temperatura
SO ₂	Dióxido de enxofre	BEN	Benzeno	VV	Velocidade do vento
NO	Monóxido de nitrogênio	TOL	Tolueno	DV	Direção do vento
NO ₂	Dióxido de nitrogênio	ERT	Enxofre reduzido total	P	Pressão atmosférica
NO _x	Óxidos de nitrogênio			RAD	Radiação Total e UVA

Figura 3 - Rede automática com estações monitoras da RMS (UGRHI 6) e variáveis medidas.

Fonte: Adaptado de (CETESB, 2021).

O mapa da Figura 4 mostra a localização das estações apresentadas na Figura 3.

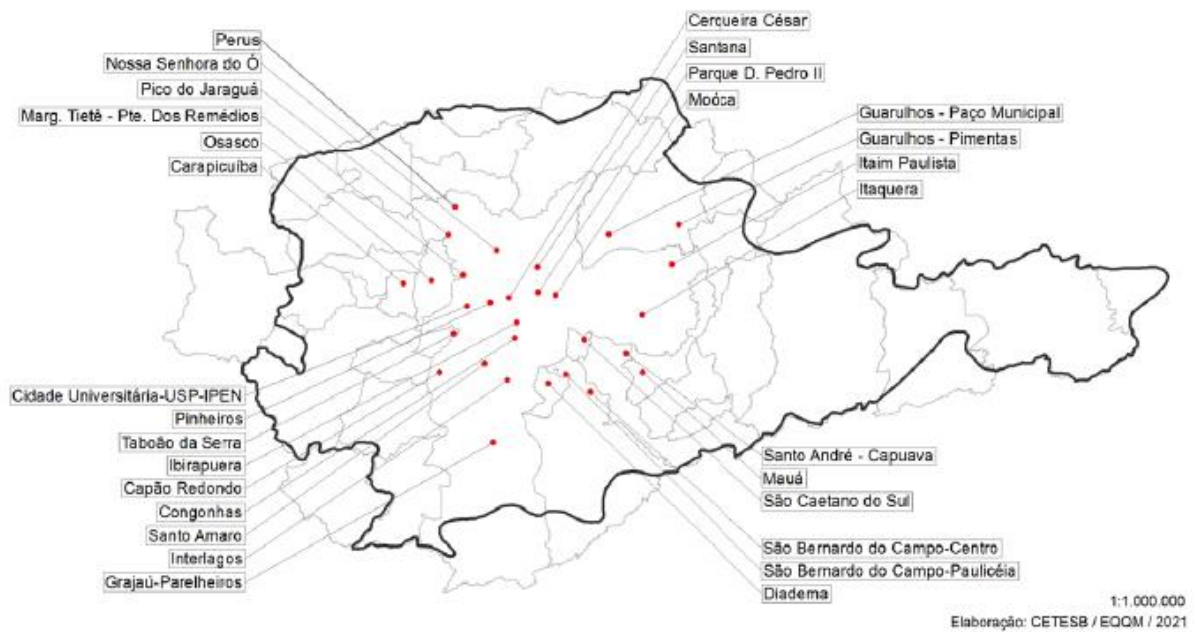


Figura 4 - Localização das estações da Rede Automática na RMSP.

Fonte: (CETESB, 2021).

1.2. Objetivos

Esta dissertação de mestrado tem por objetivo estudar a distribuição espacial de poluentes atmosféricos na Região Metropolitana de São Paulo (RMSP) em períodos específicos, assim como investigar eventuais redundâncias, através da aplicação de técnicas estatísticas de análise multivariada. Estes poluentes são monitorados nas estações da Rede Telemétrica da Companhia Ambiental do Estado de São Paulo (CETESB) a fim de determinar a qualidade do ar na RMSP.

Neste trabalho serão investigados tanto poluentes primários quanto secundários, a saber: Ozônio (O_3), partículas inaláveis (MP_{10}) e partículas inaláveis finas ($MP_{2,5}$). Estes poluentes foram selecionados por serem de maior interesse para o estudo de qualidade do ar na RMSP.

Como essa rede de monitoramento representa um investimento alto, além dos custos de manutenção, há o interesse em se estudar a possibilidade de diminuir o número de estações monitoras, ou dos analisadores presentes em cada estação, sem perda significativa de representatividade nos dados coletados. Assim, a partir dos resultados do estudo, eventuais medidores redundantes podem ser removidos ou realocados para outros locais de interesse, sem implicar custos adicionais.

1.3. Estrutura da Dissertação

A dissertação é composta de cinco capítulos. O primeiro capítulo, Introdução, apresenta a contextualização sobre a qualidade do ar e as motivações do estudo, relacionadas à otimização do monitoramento dos poluentes de maior interesse para a RMSP, a saber: O₃, MP₁₀ e MP_{2,5}.

No segundo capítulo, Revisão Bibliográfica, são explicadas as principais características da qualidade do ar da Região Metropolitana de São Paulo, assim como a aplicação dos métodos de estatística multivariada utilizados na análise das medições das concentrações dos poluentes.

No terceiro capítulo, Metodologia, são descritas as bases de dados utilizadas e as técnicas de análise multivariada para o estudo da distribuição espacial dos poluentes e identificação de correlações entre as variáveis.

No quarto capítulo, Resultados e Discussão, são apresentadas as estações e/ou analisadores que apresentaram maior correlação em suas medições, e é analisada a possibilidade de eliminação/relocação dos analisadores possivelmente redundantes.

Por fim, no quinto capítulo, Conclusões, apresenta-se os resultados do estudo, a interpretação deles, e sugestões para continuidade do trabalho em próximos estudos.

2. REVISÃO BIBLIOGRÁFICA

2.1. Qualidade do Ar na Região Metropolitana de São Paulo

A Região Metropolitana de São Paulo (RMSP) é a maior megacidade do hemisfério sul e o maior polo econômico do Brasil, compreendendo 39 cidades e totalizando uma população de cerca de 22 milhões de habitantes. Isso corresponde a 47% da população total do estado de São Paulo (CETESB, 2021).

A região é caracterizada por condições meteorológicas irregulares, com grandes variações de características em curtos períodos, apresentando um desafio para a elaboração de previsões de concentrações de poluentes. Apesar de existirem estudos (ANDRADE *et al.*, 2017; DE PAULA, 2017) que combinam de maneira bem sucedida o uso de modelagem de qualidade do ar com a mensuração das concentrações de poluentes no intuito de descrever a dinâmica de formação fotoquímica de poluentes, ainda há espaço para melhorias nos inventários de emissões da RMSP. As principais dificuldades são a falta de informações sobre fontes estacionárias e emissões evaporativas para descrever a formação de poluentes secundários na região (ANDRADE *et al.*, 2017). A complexidade climática da região também está atrelada à diversidade de suas condições. O estado de São Paulo possui, segundo a classificação climática de Köppen, oito zonas climáticas diferentes, que podem ser divididas em dois grupos: Zonas úmidas subtropicais (67%) e zonas tropicais (33%) (ALVARES *et al.*, 2013).

A quantidade expressiva de veículos na RMSP, na ordem de 8,8 milhões em 2021 segundo a Secretaria Nacional de Trânsito (SENATRAN), contribui para a deterioração da qualidade do ar na região, que ocorre, especialmente, devido às emissões veiculares. A região concentrou 48% da frota total do estado em 2019, em uma área correspondente a 3,2% do território estadual (CETESB, 2021).

A Companhia de Tecnologia de Saneamento Ambiental do Estado de São Paulo (CETESB) foi criada no final da década de 1970, dando início aos primeiros estudos de qualidade do ar na RMSP. O estado de São Paulo passava naquela época por um período de crescimento econômico e populacional. Por conta do processo de urbanização desorganizado e o aumento intensivo da quantidade de fontes de poluição atmosférica, houve o agravamento da degradação da qualidade do ar, evidenciada pelo aumento das concentrações de poluentes monitorados. Na década de 1980, as indústrias passaram a implementar programas para

controlar as emissões de dióxido de enxofre (SO₂) através da mudança de caldeiras a óleo residual para caldeiras elétricas (ANDRADE *et al.*, 2017).

Como resultado, as fontes fixas de emissão foram perdendo o protagonismo da poluição atmosférica para as fontes móveis, ou seja, as emissões veiculares passaram a ser mais preocupantes para a qualidade do ar da RMSP. Com a intenção de reduzi-las, foi implementado através da Resolução do CONAMA no. 18/1986, o Programa de Controle da Poluição do Ar por Veículos Automotores (PROCONVE), que se tornou o principal agente para a redução das emissões de poluentes nas cidades brasileiras. Foram estabelecidos os seguintes objetivos no Programa:

I - Reduzir os níveis de emissão de poluentes pelo escapamento e por evaporação, visando o atendimento aos padrões nacionais de qualidade ambiental vigentes;

II - Promover o desenvolvimento tecnológico nacional, tanto na engenharia de projeto e fabricação, como também em métodos e equipamentos para o controle de emissão de poluentes;

III - Promover a adequação dos combustíveis automotivos comercializados, para que resultem em produtos menos agressivos ao meio ambiente e à saúde pública, e que permitam a adoção de tecnologias automotivas necessárias ao atendimento do exigido por esta Resolução (CONAMA, 2002).

Este programa reduziu a emissão dos principais poluentes atmosféricos primários, como CO, NO_x, SO₂ e MP₁₀, além de poluentes secundários, como ozônio e materiais particulados finos. Segundo o portal do governo federal brasileiro (www.gov.br), com a implementação do PROCONVE/PROMOT pelo Ibama, observa-se uma redução de 98% na emissão de poluentes provenientes de fontes móveis entre a década de 1980 e 2022. Antes dos programas, a emissão média, por exemplo, de monóxido de carbono (CO) de um veículo leve era de 54g/km. Atualmente essa emissão está por volta de 0,4g/km.

Outro programa que favoreceu a redução das emissões de CO por frotas veiculares foi a criação do Programa Nacional do Etanol, conhecido como “Proálcool”, iniciado em 1975 com o objetivo de estimular o uso do etanol como combustível. Este programa, que foi possível devido à sinergia entre governo, universidades, instituições de pesquisa e empresa, impulsionou a produção de bioenergia no país desde a sua criação. Inicialmente, a principal motivação da criação do Proálcool era diminuir a dependência do petróleo no Brasil, devido ao súbito aumento de preço em 1973 no primeiro choque do petróleo. Mas logo observou-se que o etanol, além de representar vantagens econômicas, representava também uma menor

pegada de carbono quando comparado à gasolina ou outros combustíveis. A combustão de etanol de cana-de-açúcar apresenta uma redução de 61% das emissões de gases de efeito estufa (GEE) em relação à gasolina (CORTEZ, 2016).

Ao longo das décadas, apesar do crescimento constante das frotas de veículos, observou-se uma tendência decrescente da emissão de poluentes na RMSP (Figura 5). Isso se deve à utilização de veículos com novas tecnologias, menos poluidores quando comparados às antigas frotas. Em 2014, a emissão de SO₂ sofreu redução drástica, devido à alteração do teor de enxofre do diesel a partir de 2013, e da gasolina a partir de 2014. O aumento do consumo do diesel S10 em 2019, em comparação ao diesel S500, explica nova queda das emissões desse poluente (CETESB, 2021).

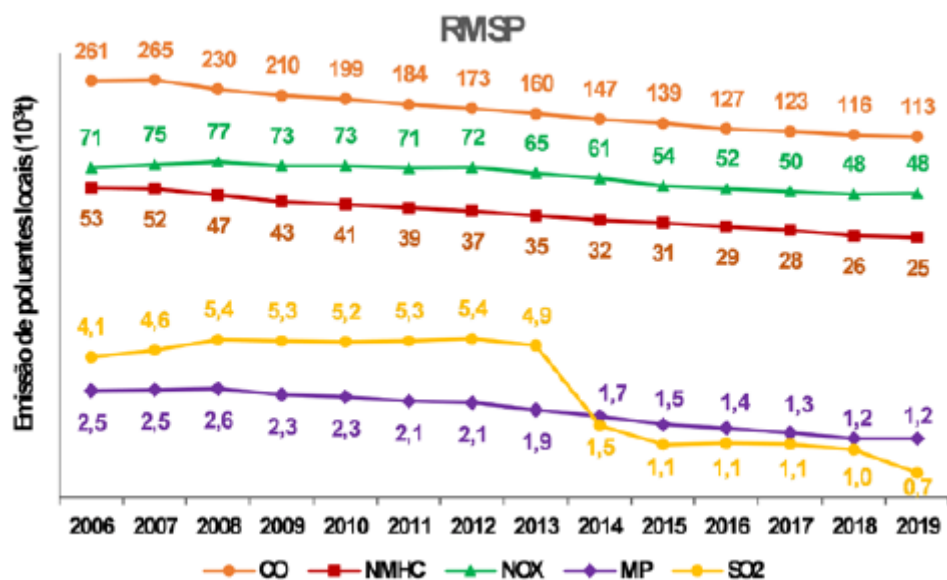


Figura 5 - Evolução das emissões de poluentes veiculares na RMSP.

Fonte: (CETESB, 2021).

As emissões atmosféricas na RMSP foram afetadas em 2020 durante a pandemia da COVID-19. Houve uma redução intensa, repentina e compulsória nas atividades produtivas, combinada à uma queda acentuada na circulação de frotas, conduzindo consequentemente à diminuição das emissões de poluentes. Entretanto, esta redução não foi suficiente para atender os valores de referência, informados pela Organização Mundial da Saúde (OMS), para alguns poluentes primários, como material particulado e NO₂. Isso reforça o desafio de manter as

concentrações de poluentes em níveis adequados para a saúde humana, especialmente nos grandes centros urbanos (CETESB, 2021).

Segundo estudo de Rudke *et al.* (RUDKE *et al.*, 2021), constatou-se que para a maioria dos poluentes não houve diminuição substancial das emissões durante o período de pandemia, quando comparado ao ano de 2019 e ao período de 2020 antes de restrições de atividades, de janeiro a março. Os poluentes que foram mais afetados durante o período de COVID-19 foram aqueles derivados de emissões veiculares: os poluentes primários CO e NO_x. Entretanto, as concentrações de ozônio aumentaram. Concluiu-se pelo estudo que fatores como o número acima da média de focos de incêndio nos meses subsequentes ao início da pandemia, juntamente com a distribuição anormal das chuvas, contribuíram para ocultar evidências de melhoria na qualidade do ar devido à redução das atividades humanas durante o período de restrições (RUDKE *et al.*, 2021).

2.2. Descrição da Rede de Monitoramento da CETESB na RMSP

Para fins de monitoramento da qualidade do ar pela CETESB, a RMSP conta em sua Rede Automática com 28 estações fixas e uma estação móvel. As estações da Rede Automática se caracterizam pela capacidade de processar dados na forma de médias horárias, nas próprias estações em tempo real. As amostragens são realizadas em intervalos de cinco segundos. É feita a transmissão dessas médias para a central de telemetria, onde elas são armazenadas em servidor de banco de dados dedicado. É realizado um processo de validação técnica periódica dos dados e, posteriormente, são disponibilizadas de hora em hora no endereço eletrônico da CETESB (CETESB, 2021).

A rede de monitoramento da CETESB foi projetada para garantir a mensuração da qualidade do ar em diversos locais de forma otimizada e a fim de atingir a uma série de objetivos. Segundo a CETESB (CETESB, 2021), as estações foram concebidas de forma a atender a necessidades de monitoramento distintas e que permitam fornecer dados, entre outros, sobre:

- os níveis mais elevados de concentração de poluentes na atmosfera esperados para a área de abrangência da rede;
- as concentrações representativas das áreas de maior densidade populacional;
- o impacto de determinadas fontes ou grupos de fontes na poluição do meio ambiente;

- os níveis médios de concentração de poluentes.

Na classificação das redes, existe o conceito de escala espacial de representatividade para um determinado poluente. Isto está relacionado à extensão da parcela de ar que cerca a estação de monitoramento e que apresenta concentrações relativamente uniformes e similares às concentrações medidas na estação. A classificação da representatividade espacial é dependente da localização da estação, considerando-se fatores como: poluente de interesse para medição, proximidade das fontes de emissão, intensidade da concentração do poluente, condições topográficas, relevo e transporte dos poluentes, além da posição da sonda de amostragem em relação à altura de captação e à proximidade de obstáculos (edificações, árvores etc.) (FRANÇA, 2019). Este conceito é relevante pois as concentrações do poluente de interesse devem estar representadas pelos resultados do monitoramento da rede numa escala espacial compatível com os objetivos do monitoramento. A Tabela 1 apresenta as escalas de representatividade estabelecidas no Decreto Estadual nº 59.113/2013 (CETESB - COMPANHIA AMBIENTAL DO ESTADO DE SÃO PAULO, 2016):

Tabela 1 - Definição de escalas de representatividade espacial.

Escala	Descrição
Microescala	Relativa à representatividade espacial de áreas de dimensão de poucos metros até 100 metros.
Média escala	Relativa à representatividade espacial de blocos de áreas urbanas (poucos quarteirões com características semelhantes) com dimensões entre 101 e 500 metros.
Escala de bairro	Relativa à representatividade espacial de áreas de bairros urbanos com atividade uniforme e dimensões entre 501 e 4.000 metros.
Escala urbana	Relativa à representatividade espacial de cidades ou regiões metropolitanas, da ordem de 4 a 50 km.

Há diferentes interpretações para a localização das estações a depender das fontes de emissão e a natureza dos poluentes.

- a) **Fontes veiculares:** Os poluentes primários que requerem mais atenção são principalmente o monóxido de carbono e o material particulado. As estações de monitoramento de microescala são relevantes sobretudo quando se implementam medidas tecnológicas de controle ou quando ocorrem mudanças significativas no fluxo de tráfego, e é necessário avaliar os efeitos dessas

mudanças. A microescala é uma ferramenta crucial quando há grupos populacionais expostos no local durante um período significativo, sendo útil para apoiar a implementação de políticas públicas específicas, uma vez que a identificação das áreas com maiores concentrações pode ser usada para esse propósito. Em vias com alta densidade de veículos movidos a óleo diesel, a medição de dióxido de enxofre (SO₂) e dióxido de nitrogênio (NO₂) pode ser necessária.

- b) Fontes de combustão – Veiculares ou industriais:** Se o poluente de interesse é o ozônio, sua maior concentração é observada em estações de monitoramento localizadas afastadas das vias de tráfego, uma vez que o ozônio é um poluente secundário que não é emitido diretamente pelas fontes. Ele é formado a partir de reações químicas entre poluentes primários, emitidos principalmente durante processos de combustão. É importante notar que os processos de combustão liberam monóxido de nitrogênio (NO), o qual reage rapidamente com o ozônio, levando a leituras pontuais de concentração de ozônio mais baixas em locais próximos às vias do que em locais afastados. Portanto, é recomendado que a escala de monitoramento do ozônio seja feita em nível de bairro ou urbano.
- c) Fontes fixas – emissões difusas e fugitivas:** Como as emissões difusas e fugitivas oriundas de fontes fixas (não efetuadas por chaminés) causam problemas localizados, nesse caso o monitoramento mais recomendado é o de microescala ou média escala.
- d) Fontes fixas – por chaminé:** As emissões de poluentes provenientes de fontes como chaminés encontram-se a uma distância considerável do solo e da zona de respiração. Os impactos mais significativos dessas emissões geralmente ocorrem distantes das fontes. Para avaliar adequadamente esses impactos, é necessário considerar a distância e a posição em relação à estação, a intensidade das emissões, a altura da chaminé, o relevo e a topografia do local, bem como aspectos meteorológicos, como a intensidade e a predominância dos ventos.

Idealmente, modelos de dispersão devem ser usados para determinar o posicionamento adequado de uma estação de monitoramento, permitindo a verificação do

maior impacto de uma fonte específica. No entanto, é preciso ter em mente que a aplicação desses modelos deve ser cuidadosa, uma vez que pode haver pouca disponibilidade de dados meteorológicos com o detalhamento necessário para os modelos (FRANÇA, 2019).

As figuras abaixo indicam a localização dos analisadores das estações para cada poluente, em escalas de cores de representatividade:

- Verde: Microescala
- Amarelo: Média escala
- Laranja: Escala de bairro
- Vermelho: Escala urbana
- Cinza: Sem analisador do referido poluente na estação indicada.

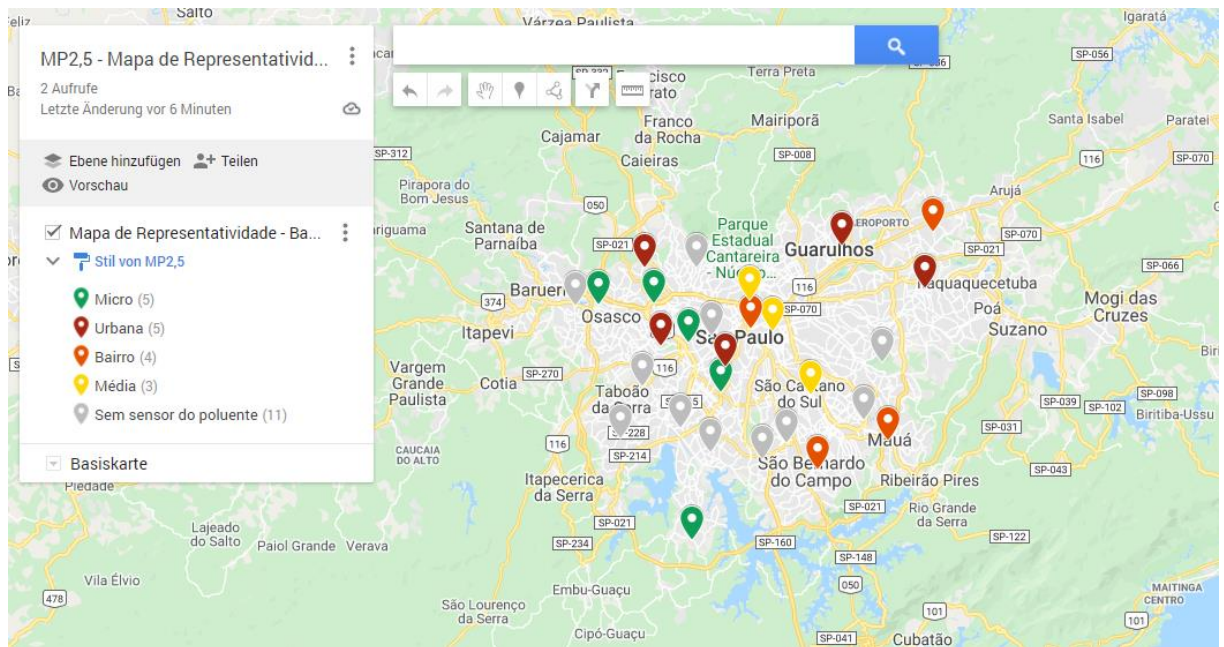


Figura 6 - Mapa de representatividade do poluente MP_{2,5} na RMSP.

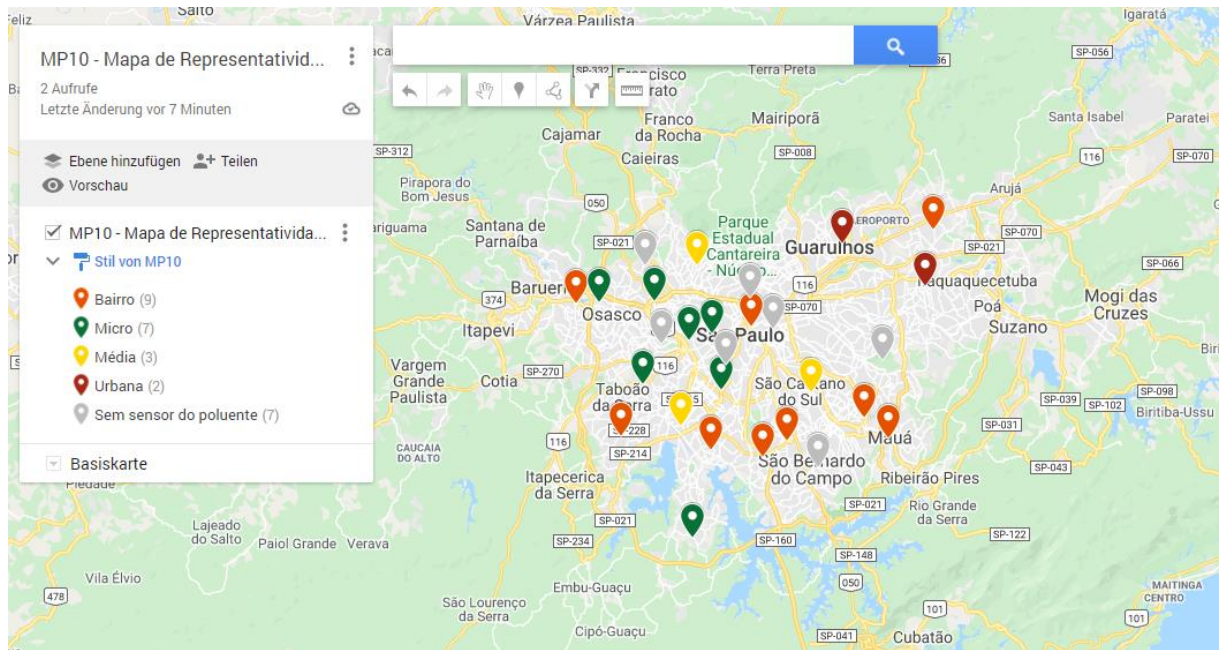


Figura 7 - Mapa de representatividade do poluente MP₁₀ na RMSP.

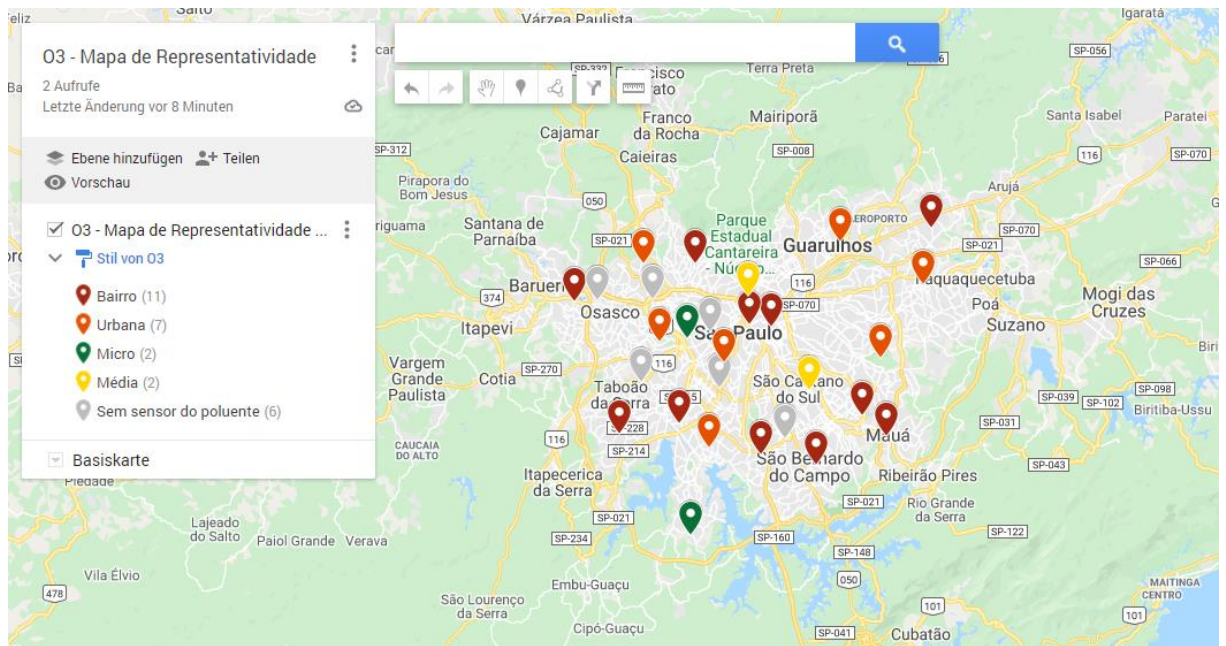


Figura 8 - Mapa de representatividade do poluente O₃ na RMSP.

Apesar da preocupação em posicionar as redes de monitoramento em localidades que otimizem a coleta de dados de concentração de poluentes da maneira mais abrangente possível, acredita-se que podem existir medições de analisadores de diferentes estações para um mesmo poluente que estão se sobrepondo. Isso pode ocorrer devido a fatores operacionais

e de planejamento, como a proximidade física entre as estações, ou à escala de representatividade, ou mesmo, as características topográficas e socioeconômicas.

Dessa forma, é de interesse da CETESB a realização deste estudo pois, caso identificadas estações e/ou analisadores com medições redundantes, é possível realocá-los sem perda de informação, com o benefício de economizar novos investimentos e custos de manutenção. Estima-se que cada analisador de medição de concentração de poluentes custa na ordem de 10-30 mil dólares, e as estações completas, 200 mil dólares.

Para avaliar se existem redundâncias, propõem-se neste estudo a aplicação de técnicas de estatística multivariada, que serão descritas a seguir.

2.3. Métodos de Estatística Multivariada

Os principais métodos de estatística multivariada aplicados para identificar redundância entre os analisadores e/ou estações de monitoramento de qualidade do ar na RMSP foram: matriz de correlação, análise de componentes principais, análise de fatores e análise de agrupamentos (*clusters*).

A matriz de correlação dos dados permite que se observem, de maneira direta, o nível de correlação entre variáveis. A técnica de análise de componentes principais permite a determinação da influência de cada variável sobre a variabilidade dos dados, a partir dos pesos das variáveis sobre os componentes principais. A análise de fatores consiste na identificação de fatores implícitos, não medidos, que podem afetar o comportamento observado do conjunto de variáveis medidas, de modo a se buscar interpretação para tais fatores. Já a técnica de análise de agrupamentos é utilizada para identificar a existência de grupos de dados ou de variáveis, visando formar subgrupos com características similares. É possível inclusive, desta forma, associar determinadas estações monitoras de forma a eliminar eventuais redundâncias de medições.

Os métodos de análise multivariada utilizados no estudo dos dados da CETESB estão descritos em mais detalhes nas próximas seções, com maior enfoque na análise de agrupamentos, que é a principal técnica aplicada neste estudo. A descrição dos métodos é baseada nas referências de Berthouex e Brown (2002), Hair Jr et. al. (1998), Himmelblau (1970), Johnson e Wichern (1998), Jolliffe (1986) e Sharma (1996).

2.3.1. Matriz de Correlação

A correlação entre duas variáveis pode ser calculada de diferentes maneiras, sendo uma das mais usuais o coeficiente de correlação de Pearson, calculado pela seguinte equação:

$$r = \frac{Cov(x,y)}{S_x S_y} = \frac{\sum_{i=1}^n (x_i - \underline{x})(y_i - \underline{y})}{\sqrt{\sum_{i=1}^n (x_i - \underline{x})^2 \sum_{i=1}^n (y_i - \underline{y})^2}}$$

Em que:

Cov(x,y): covariância entre as variáveis x e y; S_x e S_y : desvios padrões das variáveis x e y, respectivamente. \underline{x} e \underline{y} : médias das variáveis x e y, respectivamente.

Na análise multivariada, a matriz de correlação dispõe os coeficientes de correlação entre cada par de variáveis na forma de uma matriz n x n, em que n é o número de variáveis. O valor de r pode variar de -1 a +1. Quanto mais próximo de 1 for o valor absoluto desse coeficiente, mais forte é a correlação entre as variáveis. Essa matriz é de grande vantagem para análises multivariadas, pois indica de maneira rápida e fácil o quão semelhantes ou distintos são os comportamentos das variáveis quando comparadas entre si.

Para cada análise (combinação de cada poluente, ano, e período do ano), com base nos valores da matriz de correlação, determinou-se um valor mínimo a partir do qual a correlação entre duas estações foi considerada elevada, para fins de comparação tanto entre as variáveis em uma mesma análise quanto para comparar diferentes análises.

2.3.2. Análise de Componentes Principais

Dado um sistema com p variáveis, os resultados da Análise de Componentes Principais (*Principal Component Analysis*, abreviado por PCA) permitem a redução da dimensão do sistema através da geração de componentes principais, que são combinações lineares das variáveis originais obtidas a partir da matriz de covariância ou da matriz de correlação dos dados. A escolha de qual matriz utilizar afeta as combinações que resultam nos componentes principais. A matriz de correlação confere igualdade entre a variância de cada variável, permitindo que variáveis de diversas magnitudes ou variâncias muito distintas apresentem valores semelhantes. Pode-se, a partir dos resultados, selecionar um subconjunto de m novas variáveis ($m < p$, sendo p o número original de variáveis) independentes entre si,

que seja capaz de descrever, com menos variáveis, a maior parte da variabilidade dos dados e, portanto, o comportamento de um dado sistema multivariado. Além disso, o PCA permite a detecção de correlações implícitas entre variáveis ou grupos de variáveis na estrutura dos dados, que não são evidentes antes da análise multivariada.

Para obter os componentes principais é necessária a obtenção dos pesos $w_{k,j}$, coeficientes das combinações lineares das variáveis originais, sendo que k e j variam de 1 a p , para o seguinte sistema de p equações:

$$e_k = w_{k,1}X_1 + w_{k,2}X_2 + \dots + w_{k,p}X_p, \text{ sendo } 1 < k \leq p \text{ e } 1 < j \leq p$$

e_1 a e_p : componentes principais, ordenados em ordem decrescente de sua participação na variabilidade do conjunto de dados; $w_{k,j}$: pesos da variável j sobre o componente k (*component loadings*). Os pesos respeitam a condição de ortogonalidade e os p vetores são normalizados; X_1 a X_p : variáveis originais centradas na média.

A solução do problema é descrita em detalhes em livros dedicados ao tema (SHARMA, 1996). Matematicamente, o procedimento resulta na maximização da função objetivo F a seguir:

$$F = w^T \cdot Cov \cdot w - \lambda(w^T \cdot w - 1)$$

w : o vetor de pesos, e w^T é este vetor transposto; Cov : matriz de covariâncias; λ : multiplicador de Lagrange.

Trata-se de um problema característico de autovetores e autovalores, em que, para obter os pesos $w_{k,j}$, resolve-se a equação:

$$(Cov - \lambda I) \cdot w = 0$$

Como a matriz de covariância é positiva definida, todos os seus autovalores são positivos. No desenvolvimento matemático é possível provar que a variância de cada componente principal é igual a um autovalor correspondente da matriz de covariâncias. Para a obtenção dos componentes principais em ordem decrescente de importância quanto à variância dos dados, ordenam-se os autovalores da matriz Cov em ordem decrescente (em que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$) e calculam-se os autovetores correspondentes, que são os pesos dos componentes principais. A soma dos p autovalores corresponde à variância total do sistema.

Despreza-se, na seleção do subconjunto m , os componentes principais que representam menor variabilidade dos dados.

No presente estudo, para permitir uma comparação igualitária entre todas as variáveis medidas nas estações monitoras, a matriz de correlação foi utilizada para cálculo dos componentes principais. Não foi utilizada a matriz de covariância para realizar a análise de PCA, pois neste caso as diferenças de amplitude dos valores das variáveis influenciariam os resultados. Desta forma, optou-se por utilizar a matriz de correlação, com as variáveis padronizadas.

2.3.3. Análise de Fatores

A Análise de Fatores, abreviada por AF, tem alguma relação matemática com PCA, uma vez que ambas as técnicas objetivam extrair informações da matriz de correlação da base de dados estudada. Entretanto, enquanto a PCA é utilizada para simplificar o sistema em questão, buscando descrevê-lo em termos de combinações das variáveis mensuráveis, a AF possui caráter exploratório e investiga fatores implícitos que afetam a variabilidade do sistema, fatores esses que não são medidos, mas podem ser identificados pela técnica.

O valor medido de uma variável x_j ($j=1, 2, \dots, p$) de um conjunto de dados é considerado, na AF, composto por fatores comuns (F_k), entidades não mensuráveis, independentes entre si e relacionadas a todas as variáveis, e por um fator específico, ε_j , como descrito na equação a seguir:

$$x_j = \sum_{k=1}^m l_{j,k} F_k + \varepsilon_j$$

As constantes $l_{j,k}$ são os pesos dos m fatores para cada uma das p variáveis do conjunto. Na AF, encontra-se a matriz L desses coeficientes, de dimensão $(p \times m)$, a partir da qual é possível extrair informações sobre fatores implícitos não claramente explicados pelas variáveis em uma análise superficial.

Entre os resultados da AF, podem-se avaliar os efeitos dos fatores comuns e específicos, denominados, respectivamente, comunalidade e variância específica. A comunalidade corresponde à fração da variância total explicada pelos fatores comuns,

enquanto a variância específica representa a outra parcela da variância do sistema que não é devida aos fatores comuns, sendo específica da medida de cada variável. Além da obtenção dos pesos, ou seja, os elementos da matriz L , é possível realizar uma “rotação” de fatores. A rotação consiste em se aplicarem combinações lineares sobre os pesos, $l_{j,k}$, de modo a torná-los numericamente mais distantes entre si, podendo assim facilitar a interpretabilidade geral do sistema. Na literatura são descritos vários métodos de rotação, como método de rotação Varimax, método de rotação Quartimax e método de rotação Promax (HASTIE *et al.*, 2008).

2.3.4. Análise de Agrupamentos ou *Clusters*

A Análise de Agrupamentos (*cluster analysis* em inglês) é uma técnica estatística multivariada que visa a identificação de grupos de dados, ou de variáveis, e agrupá-los em grupos com alto grau de similaridade. No presente estudo, a ocorrência de similaridade entre diferentes estações monitoras pode indicar uma eventual redundância na rede de monitoramento.

Em uma amostra com n dados de um conjunto de p variáveis, X_1, X_2, \dots, X_p , a Análise de Agrupamentos tem por objetivo a identificação de g grupos, de acordo com critérios de semelhança entre os dados ou entre os grupos de dados, além de possibilitar a interpretação dos resultados do conjunto dos g grupos obtidos. Esses resultados dependem do critério de similaridade adotado no agrupamento dos dados.

As medidas de similaridade, ou de proximidade são normalmente expressas como distâncias, calculadas de diversas maneiras: distância de Minkowski, distância estatística, matriz de correlação, entre outras. Com esses critérios calculam-se por exemplo as distâncias entre as variáveis, que resultam na matriz de similaridade (matriz D). Esta matriz é a base de todos os métodos de agrupamento.

Em relação às técnicas de agrupamento, é possível classificá-las como hierárquicas, ou sequenciais, e não hierárquicas. As técnicas hierárquicas baseiam-se na ordenação de distâncias, e são iniciadas a partir da menor distância identificada na matriz D . Os dados são agrupados em sequência segundo algum critério de distância entre os grupos e, em cada etapa, forma-se um novo grupo que envolve dois ou mais grupos precedentes, diminuindo sucessivamente a dimensão da matriz de similaridade. O usuário define quantos grupos representam melhor as características do sistema em estudo. Existem também técnicas

hierárquicas divisivas para formação de grupos, em que a partir de um único grupo são realizadas divisões sucessivas segundo critérios de distância.

Pode-se avaliar o processo de formação de grupos por métodos hierárquicos através da representação gráfica, como os dendrogramas.

Existem vários critérios estatísticos que auxiliam na definição do número de grupos que deve ser mantido no final de uma análise. Dentre eles, pode-se citar:

1. Nível de similaridade: $s = 100 \cdot (1 - d_{ij}) / \max \{d_{ij}\}$
2. Valor da distância mínima entre objetos a cada atualização da matriz de similaridade.

Ao plotar e examinar o gráfico desses parâmetros a cada etapa do processo de agrupamento, podem-se identificar mudanças abruptas no comportamento, que indicam possíveis pontos de parada no processo.

No caso das técnicas não hierárquicas, define-se primeiramente o número de grupos (denominados sementes) e só então agrupam-se os dados segundo critérios de homogeneidade e de heterogeneidade entre eles. As variáveis podem “migrar” de um grupo para o outro ao longo do procedimento. Como as técnicas não hierárquicas dispensam o cálculo de uma matriz de similaridades, e não envolvem a retenção de todos os dados na memória computacional, são mais adequadas para análise de grandes quantidades de dados.

As matrizes da Tabela 2 são válidas para comparações de dispersões entre populações (grupos) e internas de cada população. \underline{X}_k representa os valores médios de cada grupo k e \underline{X} é o valor médio de todas as populações. A matriz B apresenta a soma ponderada das distâncias quadráticas entre grupos (pxp). A matriz W apresenta a soma das distâncias quadráticas internas ao grupo, isto é, entre cada observação e a média de todas as observações em cada grupo (pxp). Já a matriz SSCP, soma das duas últimas, indica a dispersão de uma população em relação ao valor médio global.

Tabela 2 - Tipos de dispersão para análises de agrupamentos.

Tipo de dispersão	Matriz de dispersão	Graus de liberdade
Entre grupos	$B = \sum_{k=1}^g n_k (\underline{X}_k - \underline{X})^T (\underline{X}_k - \underline{X})$	g-1

Interna em cada grupo	$W = \sum_{k=1}^g \sum_{j=1}^{n_k} (X_{kj} - \underline{X}_k)^T (X_{kj} - \underline{X}_k)$	$\sum_{k=1}^g n_k - g$
Totais	SSCP = B + W	$\sum_{k=1}^g n_k - 1$

Com o objetivo de identificar redundâncias entre as estações monitoras, a análise de clusters consistiu em agrupar variáveis, isto é, as próprias estações, e não os dados medidos por elas. Trata-se de uma técnica hierárquica que pode ser feita através de diferentes métodos de ligação, os quais determinam qual medida de distância entre os dados será utilizada para dar origem aos grupos. A Figura 9 apresenta os principais métodos de ligação da análise de clusters, e são eles: ligação simples (*single linkage*, também conhecido por “vizinho mais próximo”), ligação completa (*complete linkage*, também conhecido por “vizinho mais distante”), ligação média (*average linkage*) e ligação centroide (*centroid linkage*).

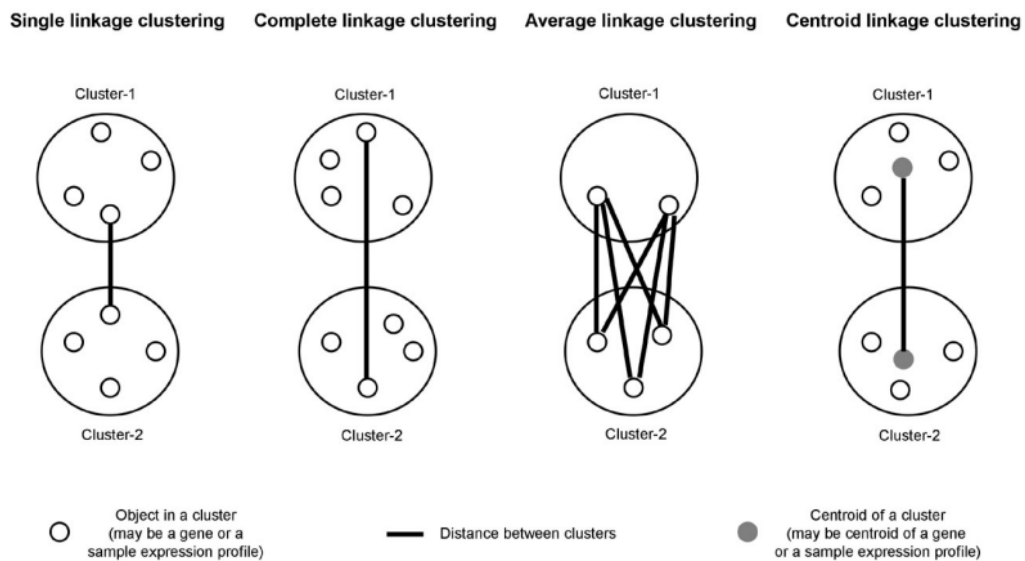


Figura 9 - Diferentes métodos de ligação para a Análise de Agrupamentos.

Fonte: (MADAN BABU, 2004).

Fujikoshi apresenta em suas publicações a utilização de múltiplas técnicas multivariadas para formular e testar hipóteses de redundância em bases de dados, assim como a comprovação matemática e estatística (FUJIKOSHI, 1989; FUJIKOSHI; KHATRI, 1990).

Há mais exemplos na comunidade científica sobre a aplicação de estatística multivariada para temáticas ambientais. Na China, utilizou-se com sucesso Análise de Componentes Principais (PCA), Análise de Cluster (CA) e análise de correlação para distinguir fontes naturais das antropogênicas para PHE's (elementos potencialmente perigosos) no solo de Beijing, uma das maiores metrópoles do mundo (YUAN *et al.*, 2013). Em Portugal, há o estudo de aplicação de análise estatística multivariada em dados hidrogeológicos de Lisboa, com o objetivo de estudar a relação entre as formações geológicas e a composição química da água subterrânea, ou seja, verificar se as espécies químicas dissolvidas na água correspondiam à composição das formações geológicas da região (SILVA *et al.*, 2002).

3. METODOLOGIA

Os dados de qualidade do ar utilizados neste estudo, coletados pela rede de monitoramento automática da CETESB na RMSP, foram fornecidos pela CETESB após tratamento prévio para excluir dados inválidos. Trata-se de médias horárias de cada variável medida por cada estação monitora, para todo o período do estudo.

A CETESB adota critérios de representatividade temporal de dados para as suas estações monitoras. Caso estes critérios não sejam atendidos para uma determinada estação ou período, conclui-se que as falhas de medição ocorridas comprometem a interpretação do resultado obtido e estes são classificados como “dados inválidos” (CETESB, 2021). Os critérios de representatividade de dados utilizados pela CETESB para a rede automática são:

- Média horária: 3/4 das medidas válidas na hora;
- Média diária: 2/3 das médias horárias válidas no dia;
- Média mensal: 2/3 das médias diárias válidas no mês;
- Média anual: 1/2 das médias diárias válidas para os quadrimestres janeiro-abril, maio-agosto e setembro-dezembro.

As bases de dados utilizadas neste estudo consistem em extrações de uma base de dados extensa fornecida pela CETESB, com medições para cada data/hora, estação de monitoramento e variável meteorológica, no período entre 01/01/2016 e 31/12/2020 (Tabela 3).

Tabela 3 - Variáveis contidas na base de dados original da CETESB.

Estação de Monitoramento
Capao Redondo
Carapicuíba
Cerqueira Cesar
Cid.Universitaria-USP-Ipen
Congonhas
Diadema
EM-50 Sesc Itaquera
Guarulhos-Paco Municipal
Guarulhos-Pimentas
Ibirapuera
Interlagos
Itaim Paulista
Marg.Tiete-Pte Remedios
Maua
Mogi das Cruzes
Mooca
N.Senhora do O
Osasco
Parelheiros
Parque D. Pedro II
Perus
Pico do Jaraguá
Pinheiros
S. Andre Capuava
S.Andre-Paco Municipal
S. Bernardo-Centro
S. Bernardo-Pauliceia
S.Caetano
Santana
Santo Amaro
Taboao da Serra

A Tabela 4 mostra, para cada variável, o significado e o tipo de cálculo utilizado para chegar ao valor numérico apresentado na base de dados.

Tabela 4 - Significado e tipo de cálculo das variáveis contidas na base de dados da CETESB (CETESB, 2021).

Sigla	Significado	Unidade	Cálculo
MP2.5	Partículas inaláveis finas	$\mu\text{g}/\text{m}^3$	Média horária
MP10	Partículas inaláveis	$\mu\text{g}/\text{m}^3$	Média horária
SO2	Dióxido de enxofre	$\mu\text{g}/\text{m}^3$	Média horária
NO	Monóxido de nitrogênio	$\mu\text{g}/\text{m}^3$	Maior média horária
NO2	Dióxido de nitrogênio	$\mu\text{g}/\text{m}^3$	Maior média horária
NOX	Óxidos de nitrogênio	$\mu\text{g}/\text{m}^3$	Maior média horária

CO	Monóxido de carbono	ppm	Maior média móvel
O3	Ozônio	µg/m ³	Maior média móvel
BEN	Benzeno	µg/m ³	Maior média horária
TOL	Tolueno	µg/m ³	Maior média horária
ERT	Enxofre reduzido total	ppb	Maior média horária
HUM	Umidade relativa do ar	%	Maior média horária
TEMP	Temperatura	°C	Maior média horária
WS	Velocidade do vento	m/s	Média horária
WD	Direção do vento	°	Média horária
PRESS	Pressão atmosférica	hPa	Média horária
RAD	Radiação Total	W/m ²	Maior média horária
UVA	Radiação UVA	W/m ²	Maior média horária

Escolheram-se como objetos de análise deste estudo os poluentes O₃, MP₁₀ e MP_{2,5}, cujas concentrações são as mais preocupantes para a qualidade do ar da RMSP. Desta forma, os dados destes três poluentes foram filtrados e subdivididos em períodos do ano: ano a ano; meses quentes e úmidos de cada ano; meses frios e secos de cada ano; período completo de 5 anos. A divisão dos dados de cada ano em períodos distintos possibilita a análise de cada poluente em diferentes condições meteorológicas, para que os resultados sejam comparados e estes efeitos avaliados quando as estações com maiores redundâncias entre si forem selecionadas.

A Tabela 5 mostra os períodos selecionados com base na análise dos dados de umidade (HUM) e temperatura (TEMP). Períodos com temperaturas médias das estações de monitoramento abaixo de 18°C foram considerados frios. Para temperaturas médias acima de 23°C foram considerados períodos quentes. Períodos com umidade relativa do ar abaixo de 70% foram considerados mais secos, e, acima de 80%, mais úmidos (APÊNDICE A).

Tabela 5 - Períodos de análise das bases de dados dos poluentes O₃, MP₁₀ e MP_{2,5}.

	Nome do Período	Código Python	Mês/Ano
1	Quente e úmido - 2016	2016qu	Janeiro/2016 - abril/2016
2	Quente e úmido - 2017	2017qu	Dezembro/2016 - março/2017
3	Quente e úmido - 2018	2018qu	Dezembro/2017 - março/2018
4	Quente e úmido - 2019	2019qu	Dezembro/2018 - março/2019
5	Quente e úmido - 2020	2020qu	Dezembro/2019 - março/2020
6	Frio e seco - 2016	2016fs	Maio/2016 - agosto/2016
7	Frio e seco - 2017	2017fs	Maio/2017 - agosto/2017
8	Frio e seco - 2018	2018fs	Maio/2018 - agosto/2018
9	Frio e seco - 2019	2019fs	Maio/2019 - agosto/2019

10	Frio e seco - 2020	2020fs	Maio/2020 - agosto/2020
11	Ano 2016	2016	Janeiro/2016 - dezembro/2016
12	Ano 2017	2017	Janeiro/2017 - dezembro/2017
13	Ano 2018	2018	Janeiro/2018 - dezembro/2018
14	Ano 2019	2019	Janeiro/2019 - dezembro/2019
15	Ano 2020	2020	Janeiro/2020 - dezembro/2020
16	Anos 2016 a 2020	2016-2020	Janeiro/2016 - dezembro/2020

Desta forma, totalizaram-se 16 bases de dados para cada poluente, com a data/horário da medição na primeira coluna e as estações nas demais, com as médias horárias das concentrações dos poluentes nas linhas.

Nas bases de dados foi feita também a seleção das estações a serem analisadas. A estação de Perus foi eliminada das análises por se tratar de uma estação de monitoramento automática móvel, utilizada em estudos temporários (CETESB, 2021).

Além disso, foram desconsideradas para a análise de redundâncias as estações de monitoramento de microescala, pois estas apresentam medições locais de concentrações dos poluentes, ou seja, específicas da área em que estão localizadas:

- a) No caso dos poluentes MP₁₀ e MP_{2,5}, que são poluentes primários, as estações de microescala estão próximas de fontes de emissão específicas (vias de tráfego). Ou seja, os analisadores destes poluentes medem as concentrações nos corredores de tráfego próximos, em um raio de poucos metros.
- b) Já no caso do ozônio, as concentrações medidas pelas estações de microescala são de ozônio proveniente de fontes de emissão afastadas da estação. Isso ocorre pois o ozônio é poluente secundário, ou seja, sua formação é resultado de reações fotoquímicas na atmosfera, na presença de luz solar e dos precursores dióxido de nitrogênio (NO₂) e compostos orgânicos voláteis (COVs). Entretanto, o monóxido de nitrogênio (NO), emitido na queima de combustíveis, consome o ozônio troposférico disponível e, com isso, as concentrações de O₃ tendem a ser baixas próximo a vias de tráfego. Assim, devido à influência do NO, é recomendável que os monitores de ozônio sejam colocados o mais afastado possível das vias de tráfego (CETESB, 2016).

Dessa forma, as estações de microescala monitoram concentrações isoladas das áreas em que estão localizadas, não sendo necessário incluí-las na análise de redundâncias entre as estações.

Eliminando-se a estação de Perus e as estações de microescala específicas de cada poluente, as bases de dados ficaram com um valor mais reduzido de colunas (estações). Para facilitar a manipulação dos dados no Python, substituiu-se os longos nomes das estações por códigos V1, V2, V3, ..., Vn, sendo n igual ao número de estações de monitoramento automáticas ativas para cada poluente. A Tabela 6 mostra a legenda do nome das estações para os poluentes analisados:

Tabela 6 - Legenda dos códigos utilizados nas bases de dados para as estações automáticas fixas da RMSP para os poluentes MP₁₀, MP_{2,5} e O₃.

	V1	V2	V3	V4	V5	V6	V7
MP10	Capao Redondo	Carapicuíba	Diadema	Guarulhos-Paco Municipal	Guarulhos-Pimentas	Ibirapuera	Interlagos
MP25	Cid.Universitaria-USP-Ipen	Guarulhos-Paco Municipal	Guarulhos-Pimentas	Ibirapuera	Itaim Paulista	Maua	Mooca
O3	Capao Redondo	Carapicuíba	Cid.Universitaria-USP-Ipen	Diadema	EM-50 Sesc Itaquera	Guarulhos-Paco Municipal	Guarulhos-Pimentas
	V8	V9	V10	V11	V12	V13	V14
MP10	Itaim Paulista	Maua	Mogi das Cruzes	Mooca	N.Senhora do O	Parque D. Pedro II	S. Andre Capuava
MP25	Parque D. Pedro II	Pico do Jaragua	S. Bernardo-Centro	S.Caetano	Santana		
O3	Ibirapuera	Interlagos	Itaim Paulista	Maua	Mogi das Cruzes	Mooca	N.Senhora do O
	V15	V16	V17	V18	V19	V20	V21
MP10	S. Bernardo-Pauliceia	S.Caetano	Santana	Santo Amaro			
MP25							
O3	Parque D. Pedro II	Pico do Jaragua	S. Andre Capuava	S. Bernardo-Centro	S.Caetano	Santana	Santo Amaro

Em algumas bases de dados para determinados períodos, havia uma quantidade significativa de dados inválidos (NaN), classificados segundo os critérios de representatividade de dados utilizados pela CETESB, apresentados na página 35. Para que estes dados não prejudicassem a aplicação das técnicas de análise estatística multivariada, para cada base de dados foram avaliadas quais estações apresentavam mais dados inválidos. Se a porcentagem de dados inválidos de uma determinada estação fosse igual ou superior a 50% do número total de dados do período, a estação era eliminada da análise.

Segue um exemplo para a base de dados do ano de 2020, para o poluente O₃. A Figura 10 mostra que a quantidade de dados NaN da estação V12 (Mogi das Cruzes) representa mais do que 8000 linhas. Como se trata de uma base de 1 ano de médias horárias, o total de linhas desta base de dados é 8760. A estação de Mogi das Cruzes está com 91% de suas medições neste ano classificadas como inválidas, pois ficou praticamente em todo o período do estudo desativada. Dessa forma, esta estação foi retirada da base antes de aplicar as análises estatísticas.

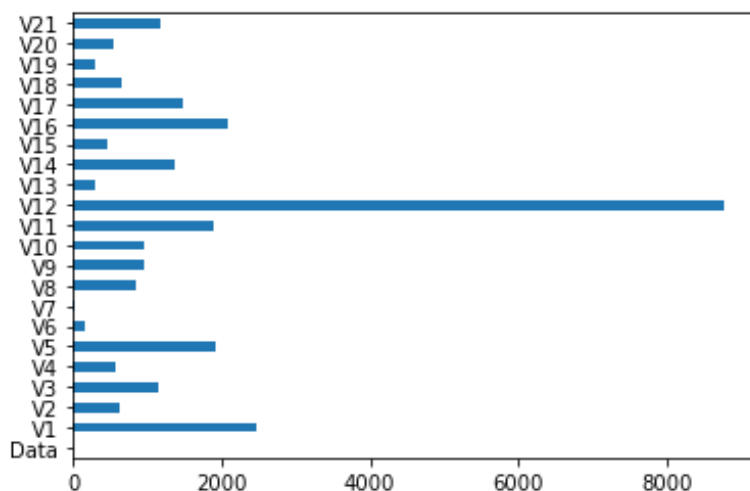


Figura 10 - Gráfico de barras da quantidade de dados inválidos para cada estação de monitoramento da base de dados do ano 2020, poluente O₃.

A Tabela 7 mostra quais estações foram eliminadas devido à quantidade de dados inválidos para cada base de dados de cada poluente.

Tabela 7 - Estações eliminadas de cada base de dados devido à alta quantidade de dados inválidos. Por período e por poluente.

Poluente	Base de Dados	Estações Eliminadas por Dados NaN
O ₃	2016qu	Mogi das Cruzes; Pico do Jaraguá; Guarulhos-Pimentas
	2017qu	Mogi das Cruzes; EM-50 Sesc Itaquera
	2018qu	Mogi das Cruzes; Diadema
	2019qu	Mogi das Cruzes; Pico do Jaraguá
	2020qu	Mogi das Cruzes
	2016fs	Mogi das Cruzes; Pico do Jaraguá
	2017fs	Mauá
	2018fs	Mogi das Cruzes
	2019fs	Mogi das Cruzes
	2020fs	Mogi das Cruzes

	2016	Mogi das Cruzes; Pico do Jaraguá
	2017	Diadema
	2018	Mogi das Cruzes
	2019	Mogi das Cruzes
	2020	Mogi das Cruzes
	2016-2020	Mogi das Cruzes
MP₁₀	2016qu	Mogi das Cruzes; Ibirapuera; Mooca; Guarulhos-Pimentas
	2017qu	Mogi das Cruzes; Ibirapuera; Mooca; Santana
	2018qu	Mogi das Cruzes; Ibirapuera; Mooca; Santana; Guarulhos-Pimentas; Diadema
	2019qu	Mogi das Cruzes; Ibirapuera; Mooca; Santana
	2020qu	Mogi das Cruzes; Ibirapuera; Mooca; Santana
	2016fs	Mogi das Cruzes; Ibirapuera; Mooca; Guarulhos-Paço Municipal
	2017fs	Mogi das Cruzes; Ibirapuera; Mooca; Santana; Guarulhos-Pimentas
	2018fs	Mogi das Cruzes; Ibirapuera; Mooca; Santana
	2019fs	Mogi das Cruzes; Ibirapuera; Mooca; Santana
	2020fs	Mogi das Cruzes; Ibirapuera; Mooca; Santana; Mauá
	2016	Mogi das Cruzes; Ibirapuera; Mooca; Guarulhos-Paço Municipal
	2017	Mogi das Cruzes; Ibirapuera; Mooca; Guarulhos-Pimentas
	2018	Mogi das Cruzes; Ibirapuera; Mooca; Santana
	2019	Mogi das Cruzes; Ibirapuera; Mooca; Santana
	2020	Mogi das Cruzes; Ibirapuera; Mooca; Santana
	2016-2020	Mogi das Cruzes; Ibirapuera; Mooca; Santana
MP_{2,5}	2016qu	Mauá; Mooca; S. Caetano; Parque D. Pedro II; Guarulhos-Paço Municipal; Santana; Guarulhos-Pimentas; Pico do Jaraguá
	2017qu	Mauá; Mooca; S. Caetano; Guarulhos-Paço Municipal; Santana; Itaim Paulista
	2018qu	Mauá; Mooca; Itaim Paulista
	2019qu	Mauá
	2020qu	Mauá; Pico do Jaraguá; Cid.Universitaria-USP-Ipen
	2016fs	Mauá; Mooca; S. Caetano; Parque D. Pedro II; Guarulhos-Paço Municipal; Santana; Pico do Jaraguá
	2017fs	Mauá; Mooca; S. Caetano
	2018fs	Mauá
	2019fs	Guarulhos-Pimentas
	2020fs	Mauá; S. Caetano; Pico do Jaraguá
	2016	Mauá; Mooca; S. Caetano; Parque D. Pedro II; Guarulhos-Paço Municipal; Santana
	2017	Mauá; Mooca; S. Caetano
	2018	Mauá
	2019	Mauá; Guarulhos-Pimentas; Itaim Paulista
	2020	Mauá; Pico do Jaraguá
	2016-2020	Mauá; Mooca; S. Caetano

Verifica-se que no caso do poluente O₃ a estação Mogi das Cruzes esteve desativada a maior parte do período de 2016 a 2020, não fazendo parte, portanto, da maior parte das

análises de verificação de similaridades entre as estações. No caso do poluente MP₁₀, as estações desativadas e, portanto, desconsideradas das análises, foram: Mogi das Cruzes, Ibirapuera, Mooca, e Santana após 2016. Já para o poluente MP_{2,5}, o ano de 2016 apresenta muitos dados inválidos, reduzindo significativamente o tamanho das bases de dados, prejudicando as análises. A estação Mauá foi desconsiderada da análise de quase todos os períodos, devido à quantidade de medições inválidas. Todas as estações com monitoramento desativado, assim como suas escalas de representatividade, podem ser consultadas no Apêndice 5 do Relatório de Qualidade do Ar no Estado de São Paulo (CETESB, 2021).

Após eliminar as estações com quantidades expressivas de dados inválidos das bases de dados, ainda sobram alguns dados NaN das demais estações. Neste caso, são eliminadas as linhas com dados inválidos antes de aplicar as técnicas de análise multivariada.

Para cada base de dados foram aplicadas as principais técnicas multivariadas com o objetivo de identificar as estações mais correlacionadas, sendo elas: matriz de correlação, análise de componentes principais, análise de fatores e análise de agrupamentos (*clusters*). Os resultados de uma análise preliminar dos dados de 2016 a 2018 estão apresentados no estudo de TCC apresentado em 2019 (PIMENTEL; FILHO, 2019).

As seções subsequentes descrevem em maiores detalhes como foram conduzidos e interpretados os métodos de análise multivariada no Python.

3.1. Matriz de Correlação

Ao avaliar a matriz de correlação, é possível identificar variáveis que possuem alta correlação, ou seja, estações com medições altamente correlacionadas. Ela apresenta coeficientes de correlação, uma medida de associação linear entre pares de variáveis, variando de -1 a 1, onde valores próximos a 1 ou -1 indicam alta correlação positiva ou negativa, respectivamente, e valores próximos a 0 indicam baixa correlação.

No Python, foi utilizado o método “.*corr()*” do Pandas para gerar a matriz de correlação, e foi possível visualizá-la de forma gráfica com uma escala de cor utilizando a função “.*heatmap()*” da biblioteca *Seaborn*, que gera um mapa de calor dos coeficientes de correlação. A Figura 11 mostra um exemplo de mapa de calor gerado a partir da função “.*heatmap()*” no Python.

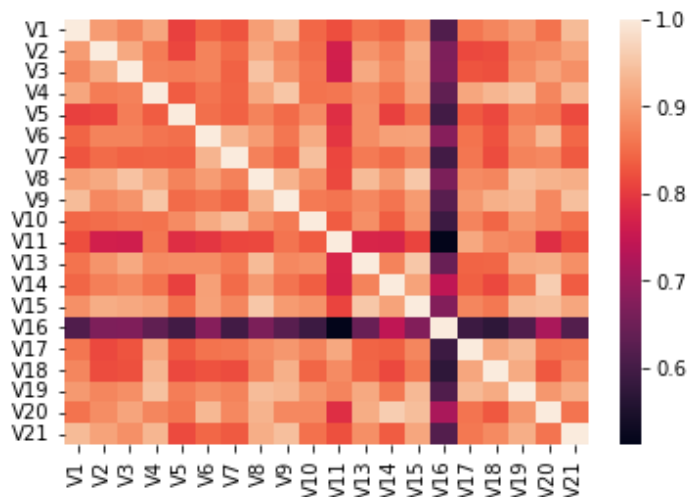


Figura 11 - Mapa de calor (*heatmap*) da matriz de correlação da base de dados de 2016 a 2020 do poluente O₃.

As cores mais claras indicam correlações mais altas entre os pares, ou seja, com coeficientes de correlação mais próximos de 1. Pares de estações nestas condições estão mais propensos a terem medições redundantes entre si.

3.2. Análise de Componentes Principais (PCA)

A análise de componentes principais (PCA) é uma técnica de redução de dimensionalidade que visa formar novas variáveis não correlacionadas (componentes principais) a partir da combinação linear das variáveis observadas originais. Estes componentes principais são ordenados em função de sua importância (maior variância).

A fim de identificar redundâncias entre as variáveis, foram feitas as seguintes análises para interpretação dos resultados de PCA:

- a) **Análise dos pesos dos componentes principais:** a PCA produz um conjunto de componentes principais ordenados por sua importância (variância). Cada componente principal é uma combinação linear das variáveis observadas e representa uma quantidade específica de variação nos dados. Variáveis com pesos semelhantes em um determinado componente principal estão mais correlacionadas entre si.
- b) **Análise do *scree plot*:** o *scree plot* é um gráfico que mostra a proporção de variação explicada por cada componente principal. Pode ser usado para determinar o número de componentes principais a serem mantidos. Se houver uma queda acentuada na proporção de variância explicada após um determinado número de componentes

principais, isso indica que as variáveis restantes são redundantes e podem ser eliminadas da análise.

A Figura 12 mostra um exemplo de *scree plot* gerado a partir de uma análise de componentes principais.

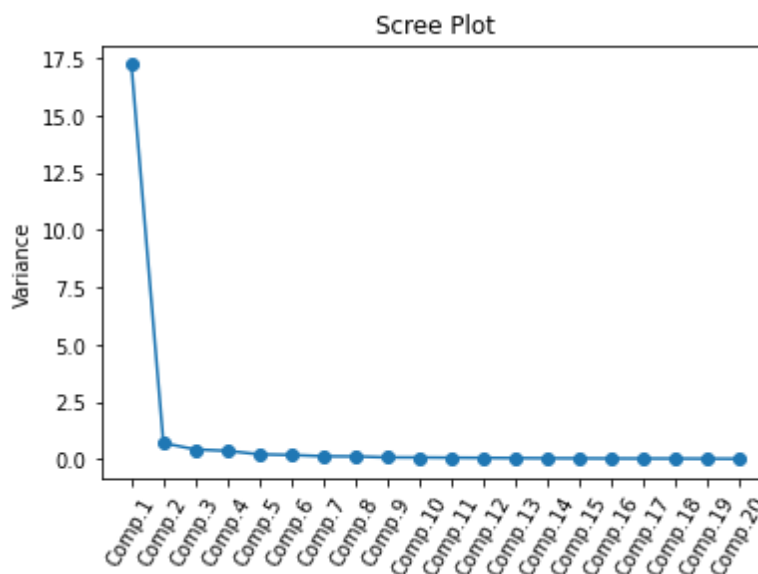


Figura 12 - *Scree plot* da análise de componentes principais da base de dados de 2016 a 2020 do poluente O₃.

O eixo Y do *scree plot* mostra a proporção da variância explicada por cada componente principal, indicados no eixo X.

Antes de aplicar a análise de componentes principais (PCA), é preciso padronizar a base de dados. A padronização é um método para escalar valores para média 0 e desvio padrão 1. Isso é útil para comparar e analisar dados que estão em diferentes escalas, enfatizando a variabilidade dos dados.

Para aplicar a análise de componentes principais (PCA) em Python foi utilizada a biblioteca *Scikit-learn*, que fornece a classe PCA para realizar análise de componentes principais. E para gerar os *scree plots* das análises, foi utilizada a biblioteca *Matplotlib*.

3.3. Análise de Fatores (AF)

A análise de fatores é uma técnica estatística que permite a identificação de fatores comuns presentes em cada medida de cada variável da base de dados. Do ponto de vista

matemático, a AF é relacionada com a PCA, mas conceitualmente são diferentes. Na AF as variáveis observadas são formuladas como combinações lineares de fatores não observados.

Antes de realizar a análise de fatores, é necessário testar se o conjunto de dados estudado permite a aplicação de análise de fatores. Existem dois métodos para verificar a adequação da amostragem:

- **Teste de esfericidade de Bartlett:** o teste de esfericidade de Bartlett verifica se as variáveis observadas se intercorrelacionam usando a matriz de correlação observada contra a matriz de identidade. Se o teste for estatisticamente insignificante, a análise fatorial não deve ser utilizada. Se o valor p for menor que um nível de significância escolhido (por exemplo, 0,05), rejeita-se a hipótese nula de que as variáveis são independentes e conclui-se que os dados são adequados para análise fatorial. Caso contrário, não se pode rejeitar a hipótese nula e conclui-se que os dados podem não ser adequados para análise fatorial.
- **Teste de adequação amostral de Kaiser-Meyer-Olkin (KMO):** o teste Kaiser-Meyer-Olkin (KMO) determina a adequação para cada variável observada e para o modelo completo. O KMO estima a proporção de variância entre todas as variáveis observadas. Os valores de KMO variam entre 0 e 1. Se o KMO for maior que 0,6, conclui-se que os dados são adequados para análise fatorial. Caso contrário, conclui-se que os dados podem não ser adequados para análise fatorial.

Em todas as análises para todos os poluentes o valor de p do teste de esfericidade de Bartlett foi aproximadamente igual a 0, confirmando que o teste foi estatisticamente significativo e as bases de dados são adequadas para a aplicação de análise de fatores. O mesmo ocorreu com o teste KMO, cujo valor para todas as análises foi maior que 0,6, tanto para as variáveis observadas quanto para o modelo completo.

Para identificar redundâncias entre as variáveis, foram analisadas as seguintes informações para interpretação dos resultados da análise de fatores:

- a) **Análise das cargas fatoriais:** a carga fatorial, ou o peso dos fatores sobre cada variável medida, é uma matriz que mostra a relação de cada variável com o fator subjacente, e corresponde à fração da variância explicada pelos fatores comuns.
- b) **Análise dos autovalores:** Os autovalores representam a variância explicada de cada fator a partir da variância total. O critério de Kaiser é uma abordagem analítica que se baseia na seleção da proporção mais significativa da variância explicada pelo

fator. O autovalor é um bom critério para determinar o número de fatores. Geralmente, um autovalor maior que 1 será considerado como critério de seleção.

- c) **Análise do *scree plot*:** a abordagem gráfica do *scree plot* para análise de fatores é baseada na representação visual dos autovalores dos fatores. Este gráfico auxilia a determinar o número de fatores onde a curva faz um “cotovelo”, que é quando há uma diminuição significativa na variância explicada pelo fator. Quanto maior a variância explicada por um fator, mais importante ele é para explicar as informações contidas nos dados.

No Python, é possível realizar o teste de esfericidade de Bartlett usando a função “.*calculate_bartlett_sphericity()*” da biblioteca *factor_analyzer*. E para calcular o teste de adequação amostral de KMO utiliza-se a função “.*calculate_kmo()*”, também da biblioteca *factor_analyzer*.

Para aplicar a análise de fatores em si, utiliza-se também a biblioteca *factor_analyzer*. Primeiro é gerado o *scree plot* com os autovalores (*eigenvalues*) no eixo Y e o número de fatores no eixo X. Em uma análise de fatores, os autovalores são uma medida da variação total explicada por cada fator extraído. No Python, os autovalores são calculados através da função “.*get_eigenvalues()*”. Neste caso, os dados também devem estar previamente padronizados.

A Figura 13 mostra um exemplo de *scree plot* gerado a partir da análise de fatores.

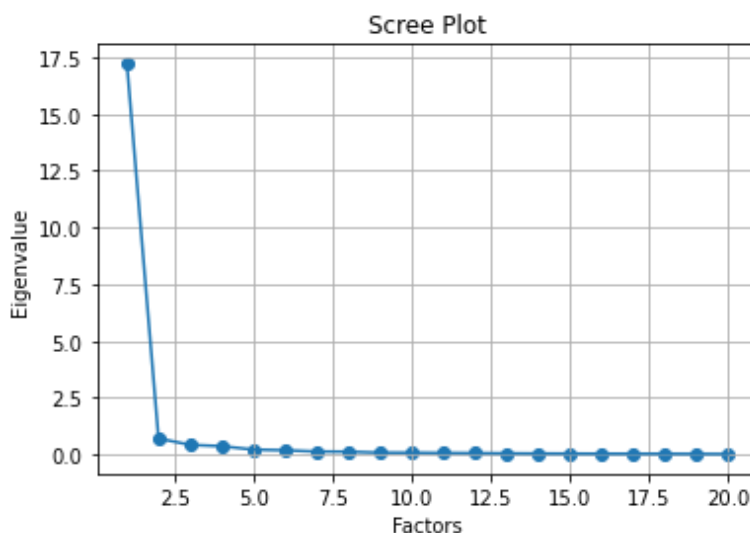


Figura 13 - *Scree plot* da análise de fatores da base de dados de 2016 a 2020 do poluente O₃.

A partir do *scree plot* é possível definir quantos fatores serão mantidos para a análise. Neste caso da Figura 13, seria apenas um fator considerado para a análise, uma vez que é o único com autovalor superior a 1.

Além de aplicar a análise de fatores, é necessário muitas vezes decidir a rotação a ser aplicada, a fim de facilitar a interpretação dos fatores extraídos e a descrição das relações entre as variáveis. Se houver apenas um fator, como no exemplo citado, não é viável realizar nenhuma rotação. Caso fosse mais de um fator, poderiam ser testados três tipos de rotação:

- Nenhuma (Python: *None*);
- Ortogonal (Python: *varimax*): que pressupõe que os fatores são independentes entre si;
- Oblíqua (Python: *oblimin*): que permite a correlação entre os fatores.

A saída da função “*fa.loadings_*” mostra as cargas fatoriais para cada variável em cada fator. Os valores mais próximos de -1 ou 1 indicam uma forte associação entre a variável e o fator correspondente. Valores próximos a 0 indicam que a variável não está fortemente associada a nenhum fator.

3.4. Análise de Agrupamentos ou *Clusters*

Neste estudo foram utilizados métodos hierárquicos de agrupamentos, uma vez que a intenção da análise é identificar as variáveis com maior grau similaridade entre elas.

Para isso, foram considerados os seguintes métodos de ligação:

- **Método de ligação simples:** Este método é indicado quando se deseja obter um agrupamento mais compacto, onde cada cluster contém os objetos mais similares. Mas pode resultar em uma grande variação de tamanho dos clusters.
- **Método de ligação centroide:** Este método é indicado quando se deseja obter clusters de tamanho semelhante, sendo útil quando se deseja identificar as variáveis que mais contribuem para a formação dos clusters.
- **Método de ligação de Ward:** Este método é indicado se a intenção é formar clusters de tamanho semelhante e que maximizem a homogeneidade dentro dos clusters. O método de ligação de Ward é útil quando se deseja identificar grupos de variáveis altamente correlacionadas, pois ele minimiza a soma dos quadrados das diferenças entre as variáveis dentro de cada cluster.

- **Método de ligação médio:** Este método é indicado se a intenção é formar clusters de tamanho semelhante e que minimizem a distância média entre as variáveis dentro de cada cluster. O método de ligação médio é útil quando se deseja identificar grupos de variáveis que estão altamente correlacionadas, mas que podem ter algumas diferenças significativas.
- **Método de ligação completa:** Este método é indicado se a intenção é formar clusters que maximizem a similaridade entre todas as variáveis em cada cluster. O método de ligação completa é útil quando se deseja identificar grupos de variáveis que estão altamente correlacionadas e que não têm muitas diferenças significativas entre si.

Os métodos de ligação simples, centroide, Ward, médio e completa foram testados em uma base de dados de cada poluente. Os dois primeiros resultaram em agrupamentos com praticamente o mesmo nível de similaridade para todas as estações, não evidenciando quais estações seriam mais similares, ou seja, sem apresentar diferenciação entre elas. Já os métodos de Ward e médio apresentaram resultados mais compatíveis com o objetivo da análise, mas não tanto quanto o método de ligação completa. Este último forneceu não apenas uma distribuição de agrupamentos visivelmente distintos, como também os resultados foram mais compatíveis com a análise da matriz de correlação. Sendo assim, a ligação completa foi utilizada em todas as análises de agrupamentos, dispostas ao longo da seção 4 deste relatório. A análise de *clusters* com o método de ligação escolhido foi aplicada em Python utilizando-se a função “*linkage*”.

Para avaliar se as observações estavam bem classificadas em seus cluster, foi calculado também o índice de silhueta médio. O índice varia de -1 a 1, e valores mais próximos de 1 indicam que a classificação foi bem feita. Valores próximos de zero indicam que as observações estão próximas das fronteiras entre clusters, e valores negativos indicam que as observações podem estar classificadas de forma incorreta. No Python foi utilizada a função “*silhouette_score*”.

Outro critério utilizado para avaliar a qualidade da análise de cluster foi a soma dos quadrados intra-cluster (*Within-Cluster Sum of Squares* ou WCSS). Trata-se de uma medida da variabilidade total dentro dos clusters: valores menores indicam que as observações dentro de cada cluster estão mais próximas umas das outras e que os clusters são mais homogêneos. No Python foi utilizada a função “*KMeans*” e “*.inertia_*”.

Uma maneira de definir o método de ligação ideal para a base de dados é comparar os resultados do índice de silhueta e da soma dos quadrados intra-cluster para diferentes métodos de ligação. O ideal é escolher o método de ligação que produz o índice de silhueta mais alto e a WCSS mais baixa, indicando uma melhor separação entre clusters e maior homogeneidade dentro de cada cluster. No caso das bases de dados deste estudo, notou-se que os cálculos de índice de silhueta médio e WCSS eram os mesmos, mesmo com a alteração do método de ligação. Isso pode indicar que a estrutura dos dados não varia muito entre os clusters formados por diferentes métodos, ou que o impacto do método de ligação não é tão significativo nesses casos.

Para observar as formações dos clusters de maneira gráfica, foram gerados dendrogramas para as análises de agrupamentos. Trata-se de gráficos que ilustram as relações de similaridade ou dissimilaridade entre os elementos de um conjunto de dados. Em um dendrograma, cada elemento é representado por uma linha ou retângulo e é agrupado com outros elementos que apresentam características semelhantes. A medida da similaridade entre esses elementos é representada pela altura da junção dos retângulos que os representam.

O dendrograma é construído a partir do algoritmo de agrupamento, que calcula a distância entre os elementos (eixo Y) e os agrupa de forma hierárquica, formando grupos com as variáveis (eixo X) cada vez maiores até que todos os elementos sejam agrupados em um único grupo.

No caso das bases de dados deste estudo, as estações de monitoramento que se “unem” em retângulos mais baixos são as mais semelhantes e, portanto, mais propensas a terem medições redundantes. A Figura 14 mostra um exemplo de dendrograma.

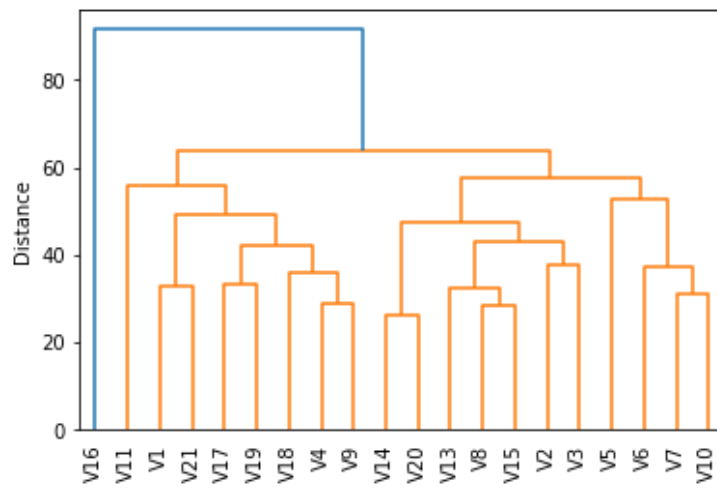


Figura 14 - Dendrograma da análise de agrupamentos da base de dados de 2016 a 2020 do poluente O₃.

4. RESULTADOS E DISCUSSÃO

As análises descritas nos itens 3.1 a 3.4 foram aplicadas a todas as bases de dados da Tabela 5, página 37, para cada poluente: O₃, MP₁₀ e MP_{2,5}.

Nas próximas subseções, estão apresentados os resultados para cada poluente nos períodos em que foram estudados.

4.1. O₃

O ozônio se forma na atmosfera através de reações fotoquímicas. Sendo assim, períodos do ano com maior incidência de radiação solar tendem a formar maiores concentrações deste poluente. Desta forma, meses quentes são de grande interesse para as análises, sobretudo os secos, uma vez que as nuvens dificultam a formação desse poluente na atmosfera. Além disso, a concentração dos precursores do ozônio, NO₂ e COVs, depende de fatores como: a taxa de emissão, que é oriunda de fontes móveis como veículos; a dispersão provocada pela instabilidade atmosférica e pela velocidade e direção do vento (DE PAULA, 2017).

Observou-se que as correlações entre as estações para este poluente são maiores do que para as medições de MP₁₀ e MP_{2,5}, da ordem de 0,94. Sendo assim, os agrupamentos formados para as estações monitoras desse poluente possuem nível de similaridade superior aos outros. Uma tendência observada foi o agrupamento de estações com maior proximidade geográfica. Por exemplo, estações localizadas na Zona Sul da RMSP agrupam-se entre si, e as da Zona Norte e Centro também.

Foram gerados e analisados os mapas de distribuições das concentrações médias e máximas de ozônio em cada período, para cada estação de monitoramento. Com estes mapas foi possível avaliar a evolução das concentrações de ozônio, assim como identificar as áreas da RMSP em que as estações captaram as maiores concentrações deste poluente.

Entre a Figura 15 e Figura 19 está apresentada a distribuição espacial da concentração média anual de ozônio para vários anos.

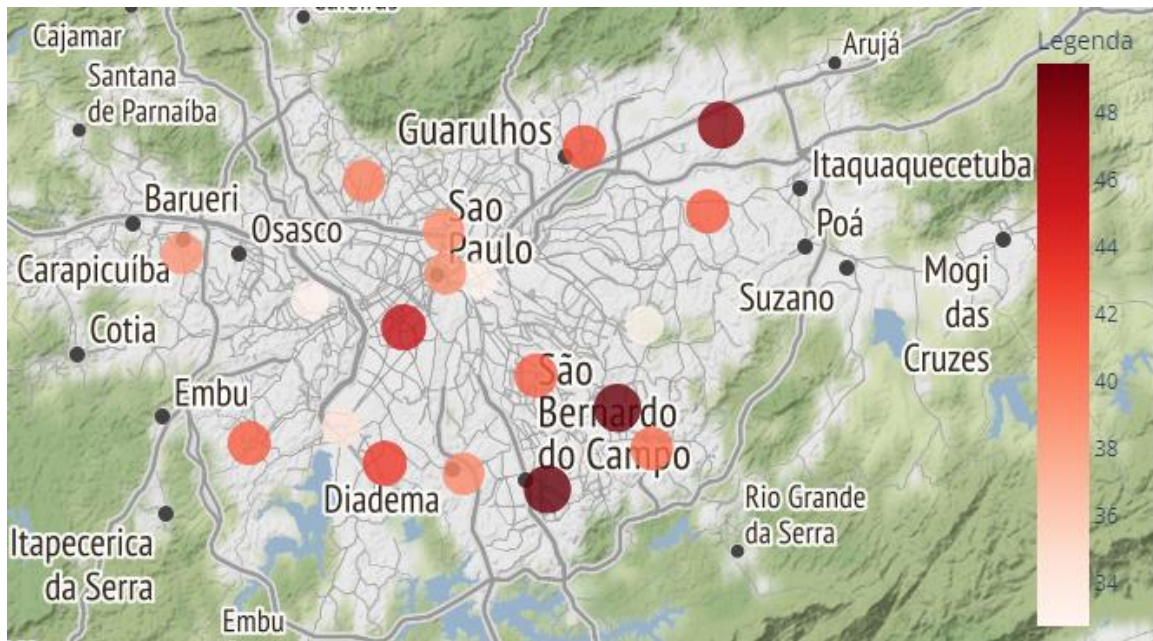


Figura 15 - Concentração média anual de O₃, em µg/m³, de 2016.

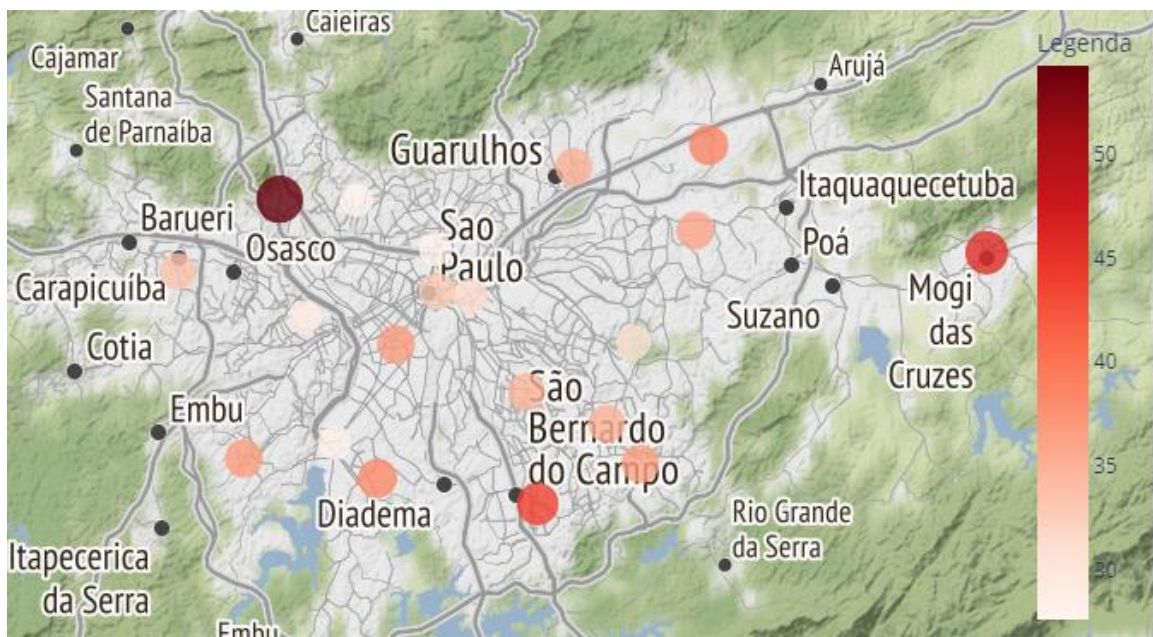


Figura 16 - Concentração média anual de O₃, em µg/m³, de 2017.

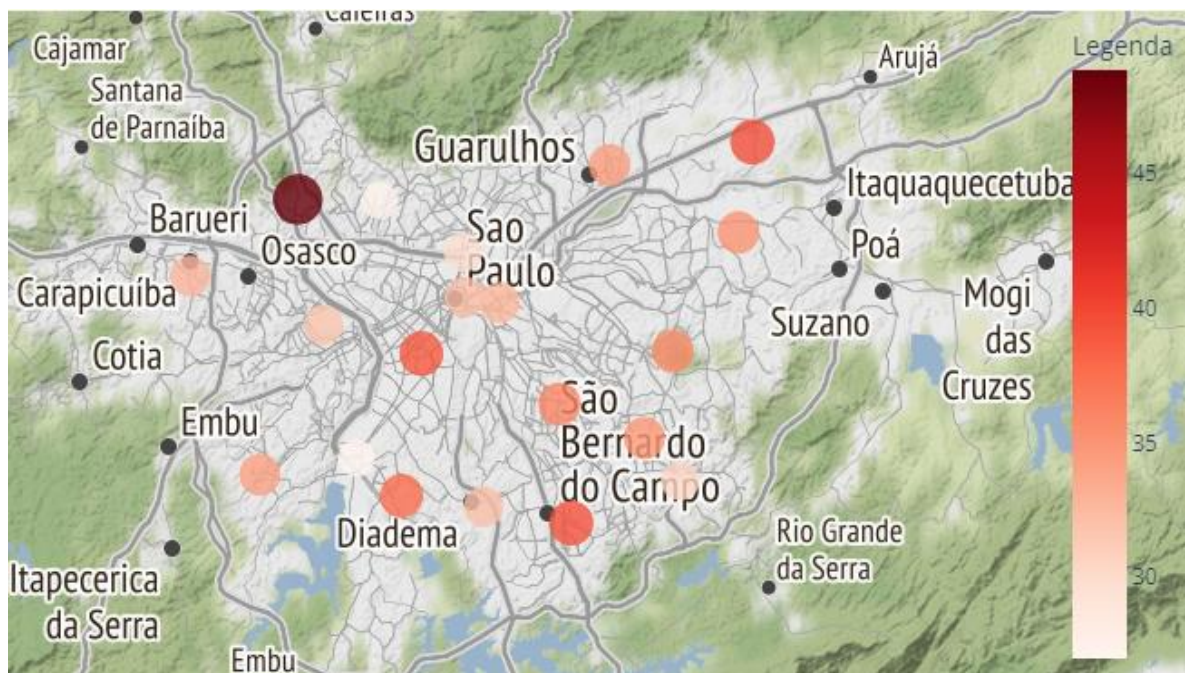


Figura 17 -Concentração média anual de O₃, em µg/m³, de 2018.

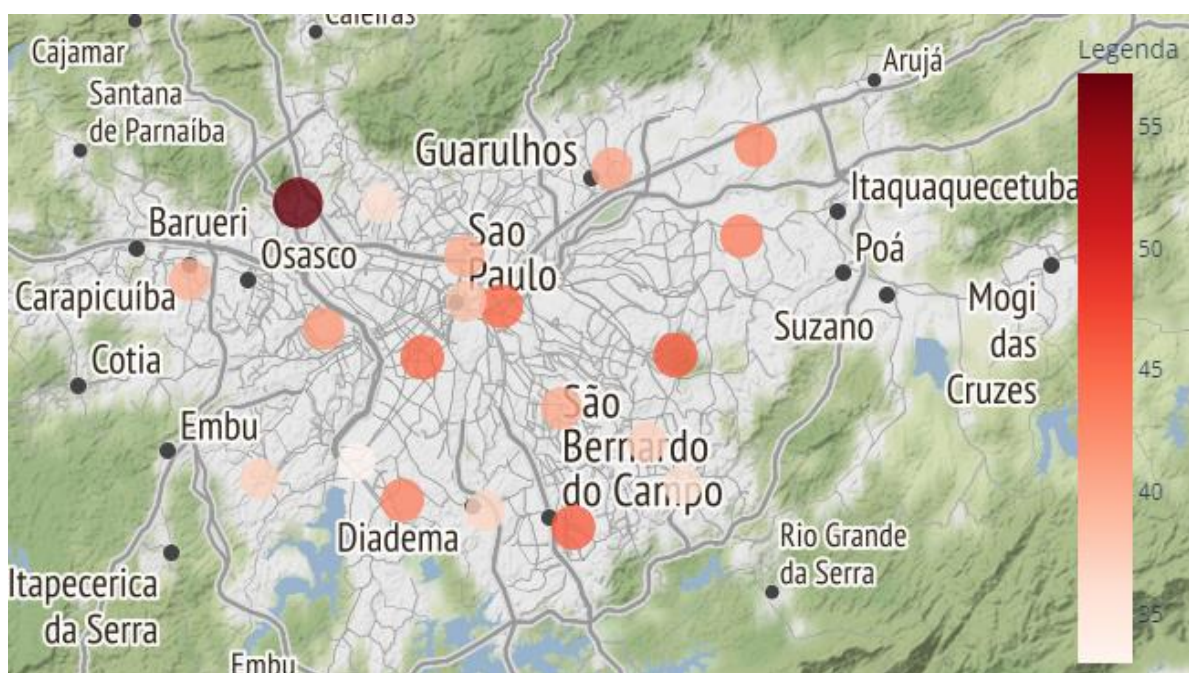


Figura 18 - Concentração média anual de O₃, em µg/m³, de 2019.

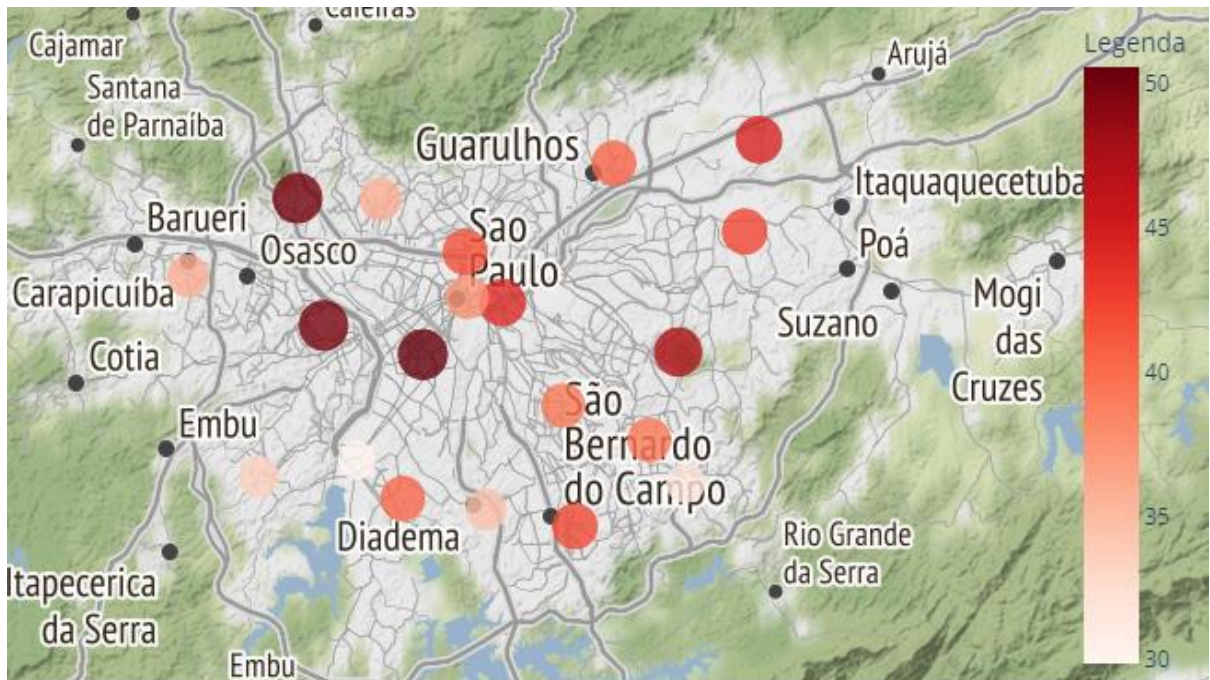


Figura 19 - Concentração média anual de O₃, em µg/m³, de 2020.

Já entre a Figura 20 e a Figura 24 estão as concentrações médias deste poluente no período quente e úmido.

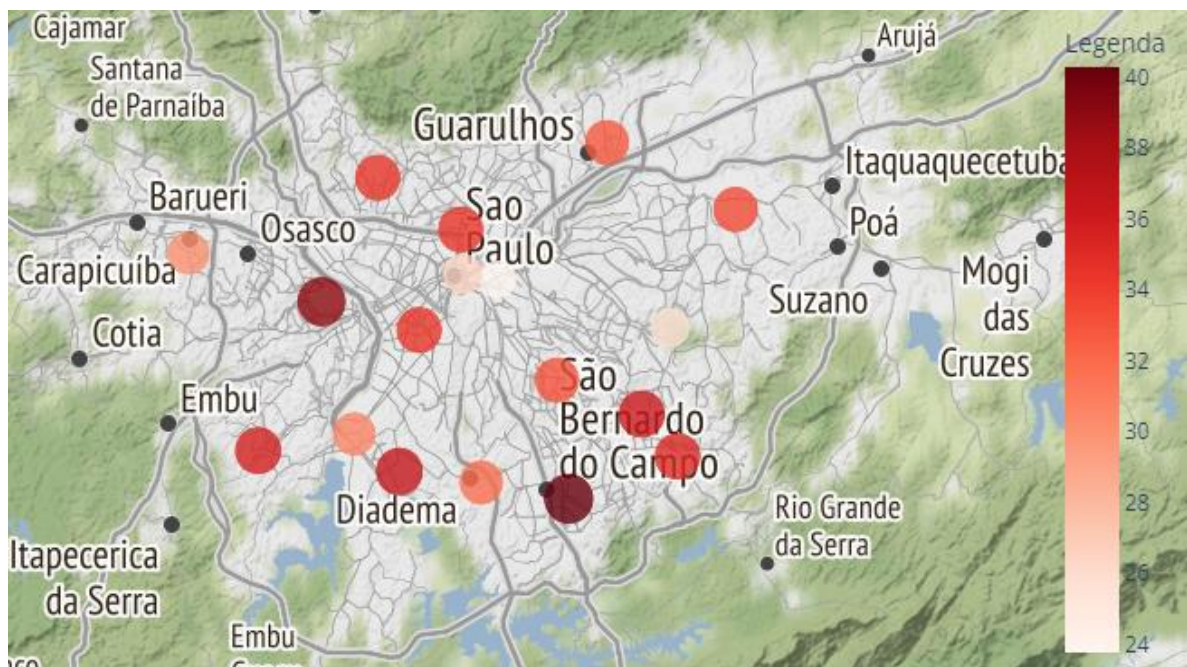


Figura 20 - Concentração média de O₃, em µg/m³, no período quente e úmido em 2016.

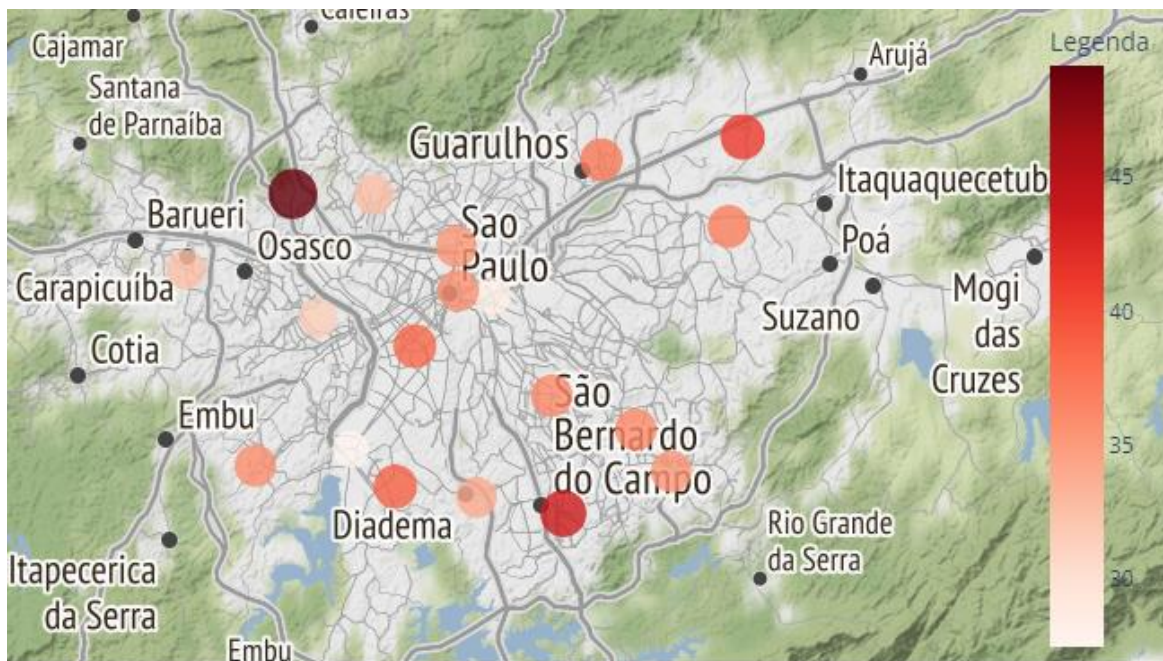


Figura 21 - Concentração média de O₃, em µg/m³, no período quente e úmido em 2017.

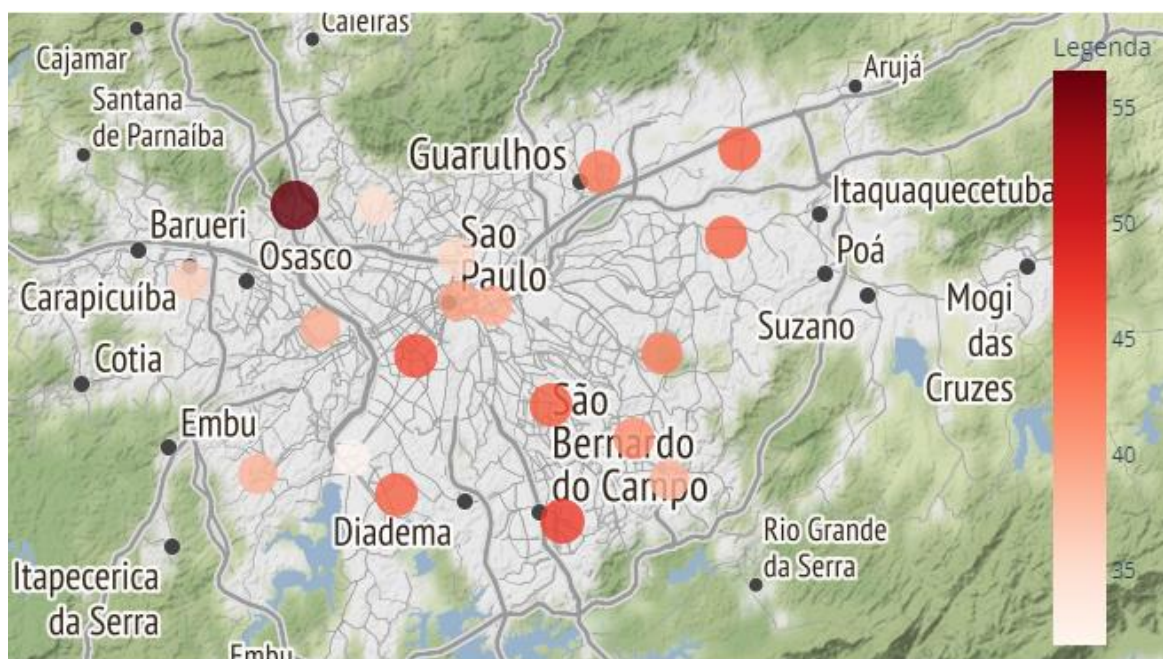


Figura 22 - Concentração média de O₃, em µg/m³, no período quente e úmido em 2018.

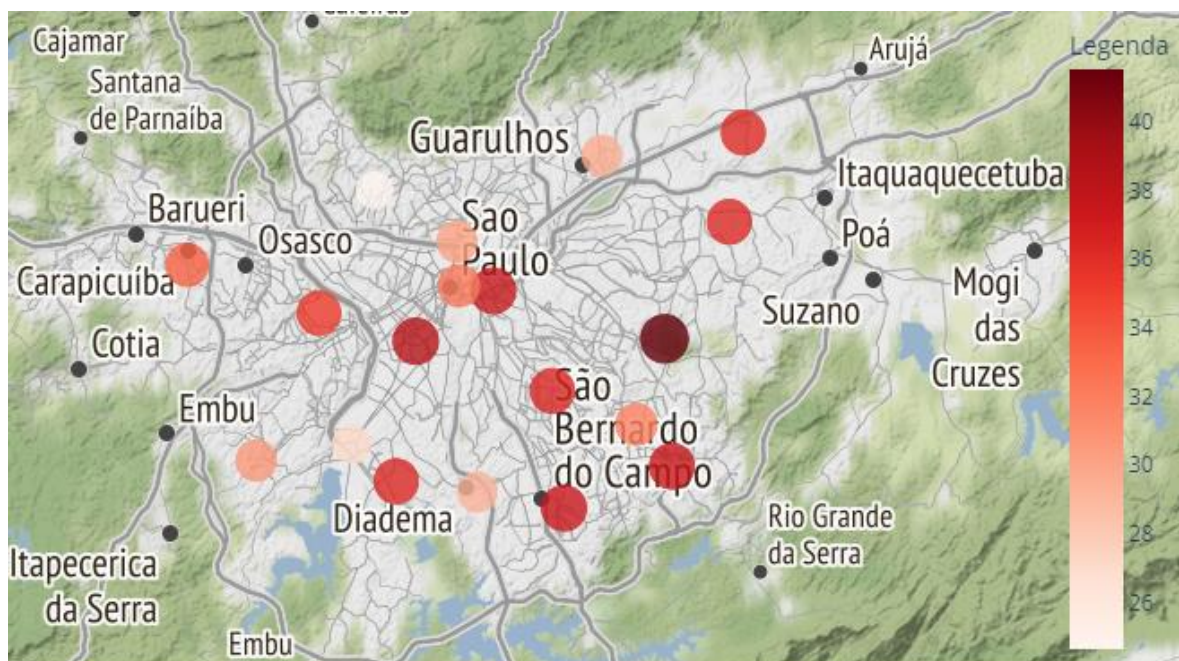


Figura 23 - Concentração média de O₃, em µg/m³, no período quente e úmido em 2019.

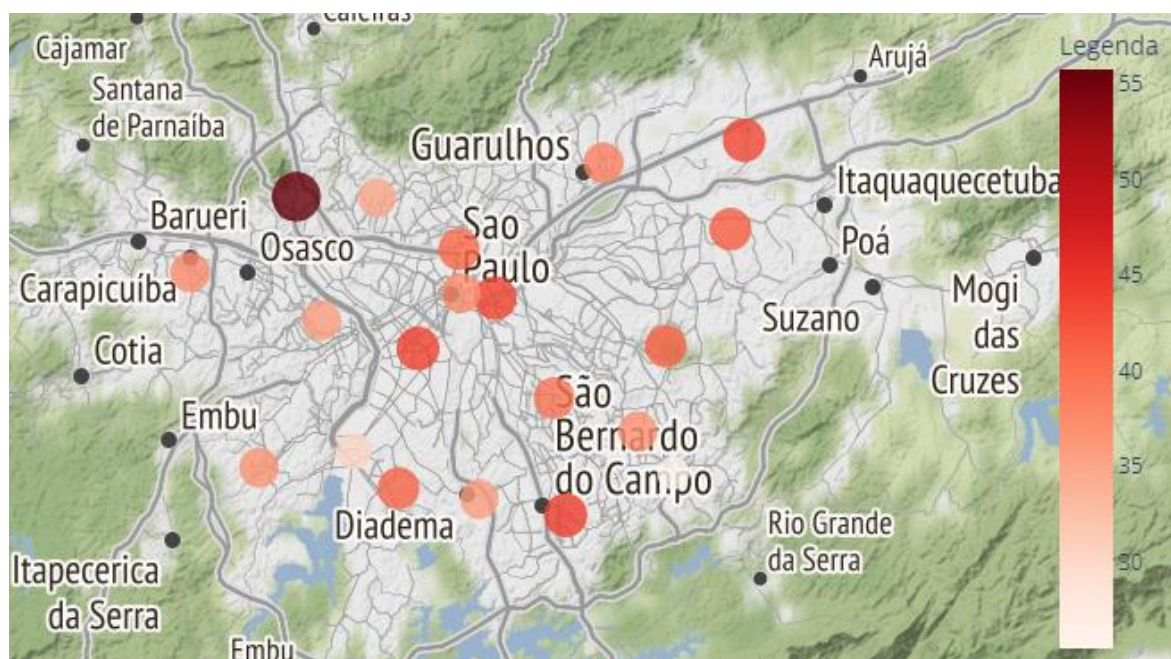


Figura 24 - Concentração média de O₃, em µg/m³, no período quente e úmido em 2020.

Já entre a Figura 25 e a Figura 29 estão as concentrações médias deste poluente no período frio e seco.

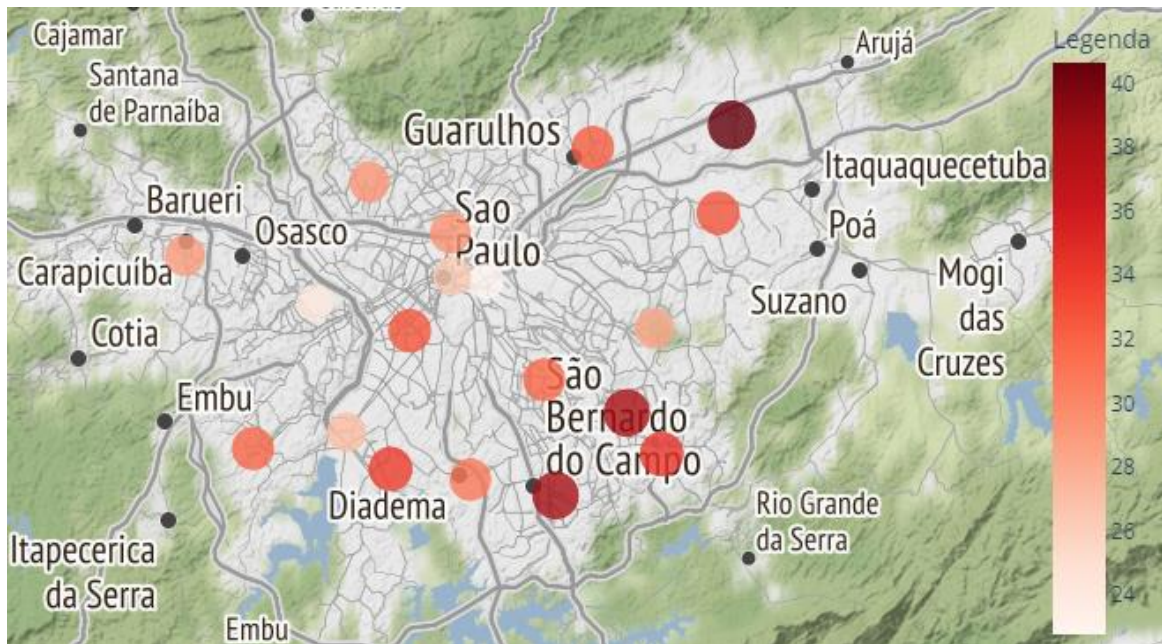


Figura 25 - Concentração média de O₃, em µg/m³, no período frio e seco em 2016.

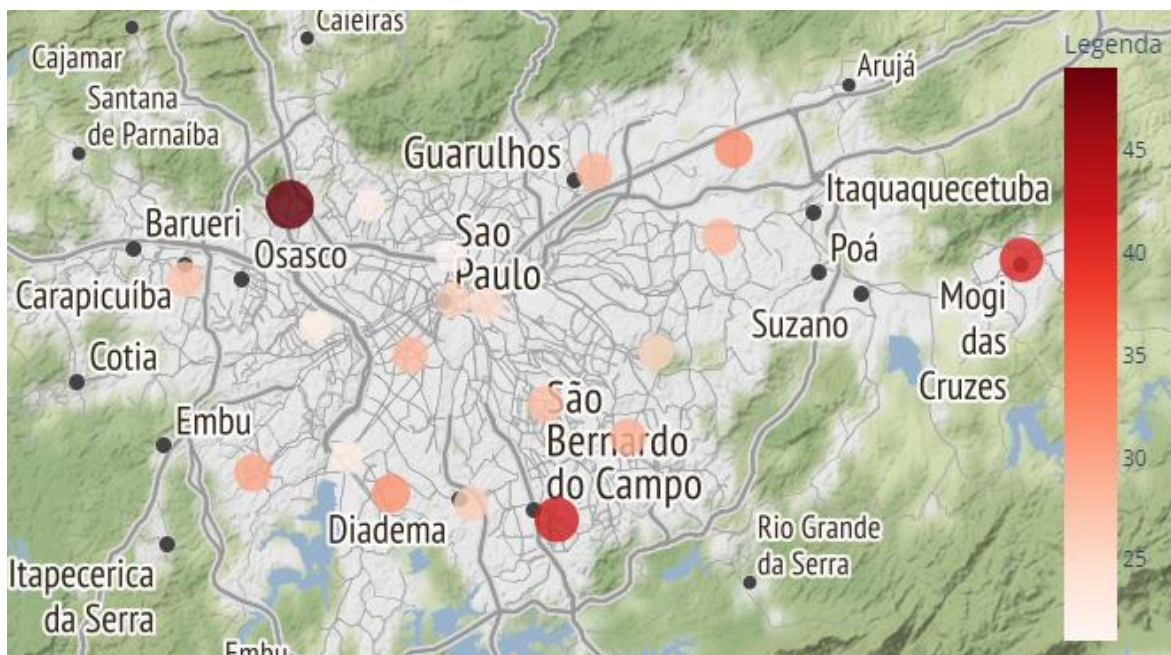


Figura 26 - Concentração média de O₃, em µg/m³, no período frio e seco em 2017.

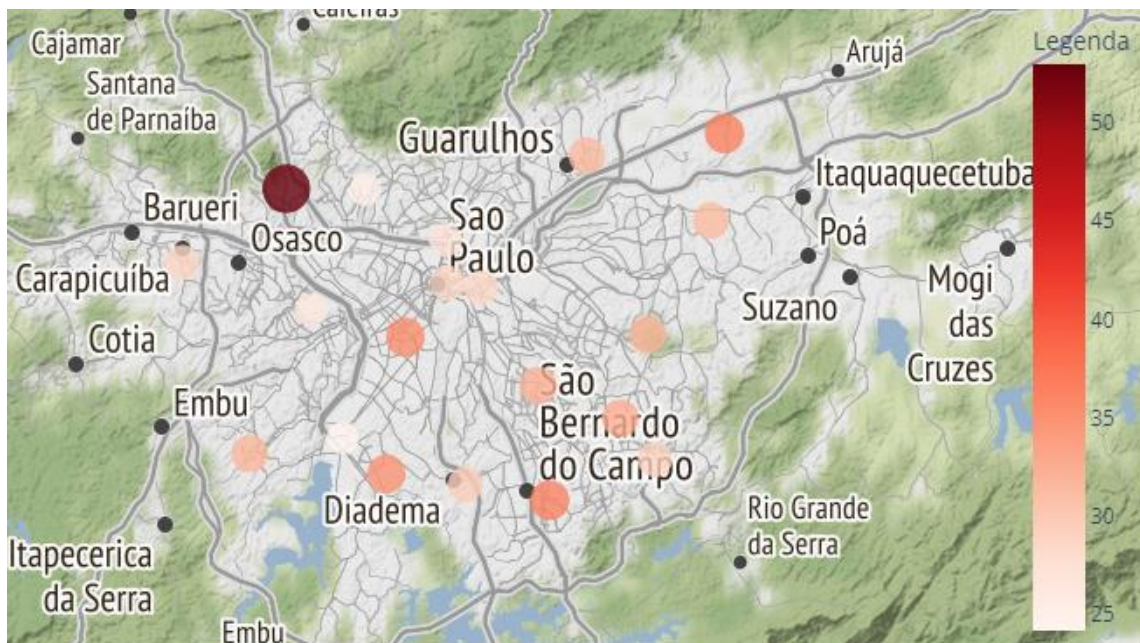


Figura 27 - Concentração média de O₃, em µg/m³, no período frio e seco em 2018.

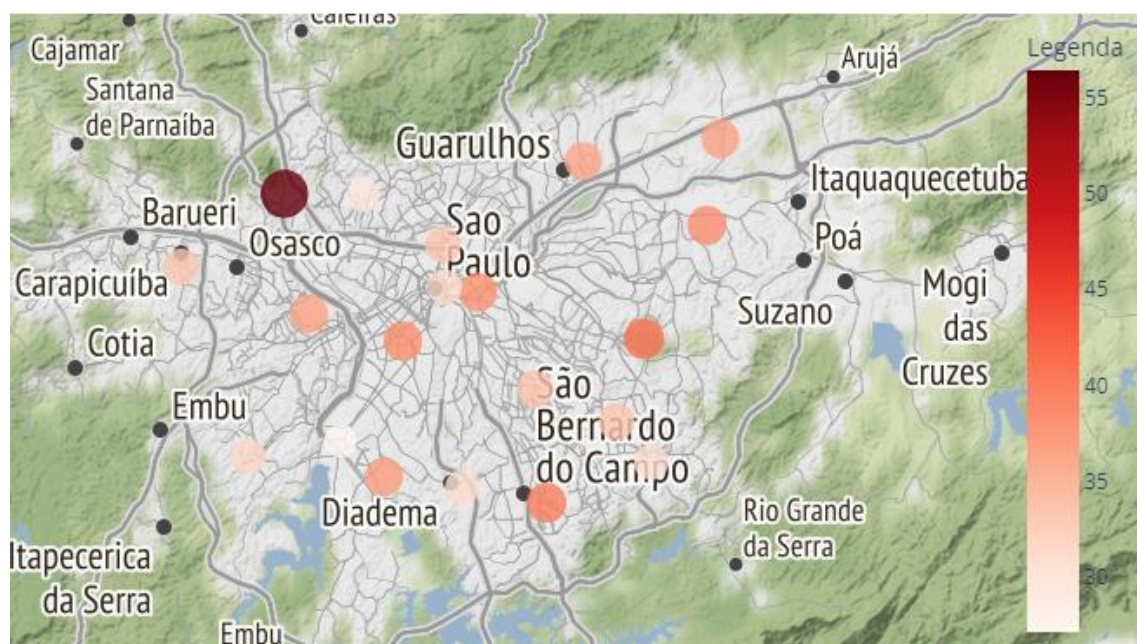


Figura 28 - Concentração média de O₃, em µg/m³, no período frio e seco em 2019.

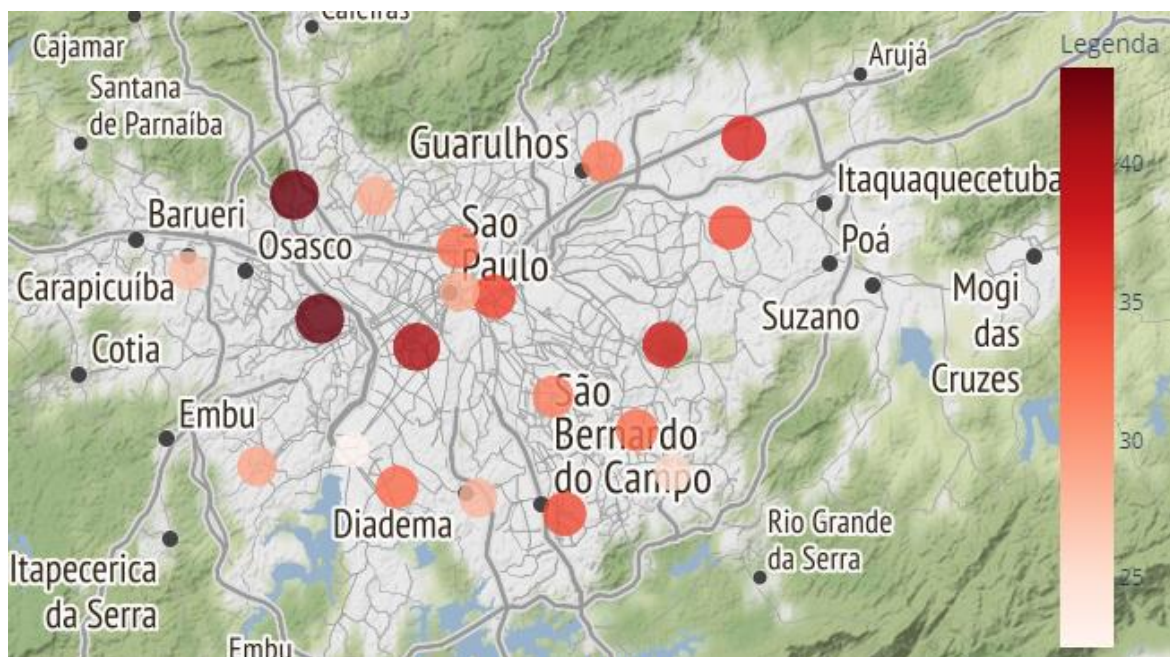


Figura 29 - Concentração média de O₃, em µg/m³, no período frio e seco em 2020.

Nota-se pelos mapas de distribuição espacial das concentrações médias de ozônio que os maiores valores se concentram nas estações mais afastadas do centro, e são detectadas principalmente pela estação Pico do Jaraguá, cuja escala de representatividade é urbana.

As estações Mooca (bairro) e Parque D. Pedro II (bairro), distantes 2,7 km uma da outra, apresentam concentrações semelhantes para quase todos os períodos analisados. As semelhanças entre as medições das estações se devem à incidência solar, que é um dos principais fatores de influência na geração de ozônio.

A Tabela 8 apresenta os resultados das análises multivariadas ano a ano com alguns dos resultados, da seguinte forma: a coluna matriz de correlação (MC) mostra os pares de estações com coeficiente de correlação acima de 0,95. No caso da análise de componentes principais (PCA), a porcentagem indica a variância explicada pelo primeiro componente principal (PC1) e as estações com pesos no PC1 superiores a 0,23. A coluna da análise de fatores (AF) mostra quantos fatores foram mantidos para análise e as estações com *loadings* acima de 0,95 no primeiro fator, que explica a maior parte da variância dos dados. E, por fim, a coluna de análise de clusters (AC) contém os 4 pares de estações com maior grau de similaridade.

Tabela 8 - Resumo dos principais resultados das técnicas de análise estatística multivariada para o poluente O₃.

Base de Dados	MC	PCA	AF	AC
2016	N. Senhora do Ó - Santana			
	Cid. Universitária-USP-Ipen - Ibirapuera			
	Ibirapuera - Parque D. Pedro II	88,4%	1 fator	N. Senhora do Ó – Santana
	Mooca - Parque D. Pedro II	Ibirapuera S. Caetano Parque D. Pedro II	Cid. Universitária-USP-Ipen Ibirapuera Parque D. Pedro II	Cid. Universitária - USP-Ipen - Ibirapuera
	Diadema - Interlagos	Diadema Interlagos Santana	Parque D. Pedro II S. Caetano Interlagos	Diadema - Interlagos
	Parque D. Pedro II - Santana	Diadema Mooca	Interlagos	Mooca - Parque D. Pedro II
	Diadema - S. Caetano			
	Carapicuíba - Cid. Universitária-USP-Ipen			
2016qu	Mooca - Parque D. Pedro II	89,7%		
	Cid. Universitária-USP-Ipen - Ibirapuera	Parque D. Pedro II Ibirapuera Cid. Universitária-USP-Ipen	1 fator Cid. Universitária-USP-Ipen Ibirapuera Parque D. Pedro II	Mooca - Parque D. Pedro II Cid. Universitária-USP-Ipen - Ibirapuera
	Interlagos - Santo Amaro	S. Caetano Santo Amaro	Parque D. Pedro II S. Caetano Diadema EM-50 Sesc Itaquera	Interlagos - Santo Amaro
	N. Senhora do Ó - Santana	Interlagos Mooca Diadema Santana	S. Caetano Diadema EM-50 Sesc Itaquera	N. Senhora do Ó - Santana
	Ibirapuera - Parque D. Pedro II	EM-50 Sesc Itaquera	Interlagos Mooca Santana Santo Amaro	
	Diadema - Interlagos	S. André Capuava Guarulhos-Paço Municipal	Mooca Santana Santo Amaro	
	Parque D. Pedro II - Santana	N. Senhora do Ó		

	Capão Redondo - Santo Amaro			
	Cid. Universitária-USP-Ipen - Parque D. Pedro II			
	EM-50 Sesc Itaquera - S. Caetano			
		88,4%		
2016fs	Cid. Universitária-USP-Ipen - Ibirapuera	Cid. Universitária-USP-Ipen Ibirapuera	1 fator	Cid. Universitária-USP-Ipen - Ibirapuera
	N. Senhora do Ó - Santana	Parque D. Pedro II S. Caetano	Cid. Universitária-USP-Ipen Ibirapuera	N. Senhora do Ó - Santana
		Interlagos Itaim Paulista Santo Amaro	Parque D. Pedro II S. Caetano	Mooca - Parque D. Pedro II
		Carapicuíba Diadema Mooca Capão Redondo	Interlagos	Interlagos - Santo Amaro
2017	Mooca - Parque D. Pedro II			
	Cid. Universitária-USP-Ipen - Ibirapuera		1 fator	Mooca - Parque D. Pedro II
	N. Senhora do Ó - Santana	87,8%	Cid. Universitária-USP-Ipen Ibirapuera	Cid. Universitária-USP-Ipen - Ibirapuera
	Mauá - S. André Capuava	Mooca Ibirapuera	Mooca Parque D. Pedro II	
	Guarulhos-Pimentas - Itaim Paulista	Cid. Universitária-USP-Ipen Parque D. Pedro II S. Caetano	Parque D. Pedro II Capão Redondo Interlagos S. André	Guarulhos-Pimentas - Itaim Paulista
	Ibirapuera - Parque D. Pedro II		Capuava Santana	Mauá - S. André Capuava
S. André Capuava - S. Caetano				

	Ibirapuera – Mooca			
	Carapicuíba - Cid. Universitária- USP-Ipen Mooca - Parque D. Pedro II			
	Mauá - S. André Capuava			
	Ibirapuera - Mooca	86,7%		
	Cid. Universitária- USP-Ipen - Ibirapuera	Ibirapuera Mooca Parque D. Pedro II	1 fator	Mooca - Parque D. Pedro II
	Ibirapuera - Parque D. Pedro II	Cid. Universitária- USP-Ipen	Cid. Universitária- USP-Ipen	Mauá - S. André Capuava
2017qu	N. Senhora do Ó - Santana	USP-Ipen S. Caetano Santana	Ibirapuera Mooca Parque D. Pedro II	Guarulhos- Pimentas - Itaim Paulista
	Guarulhos- Pimentas - Itaim Paulista	S. André Capuava Santo Amaro Diadema Interlagos	S. Caetano	Cid. Universitária- USP-Ipen - Ibirapuera
	Diadema - Interlagos			
	S. André Capuava - S. Caetano			
	Diadema - S. Bernardo-Centro			
	Mooca - Parque D. Pedro II	85,2%	1 fator	Mooca - Parque D. Pedro II
	Cid. Universitária- USP-Ipen - Ibirapuera	Cid. Universitária- USP-Ipen S. Caetano Parque D. Pedro II	Cid. Universitária- USP-Ipen Ibirapuera Mooca Parque D. Pedro II	Cid. Universitária- USP-Ipen - Ibirapuera
2017fs	Carapicuíba – Cid. Universitária- USP-Ipen	Ibirapuera Mooca Interlagos Capão Redondo	Capão Redondo Interlagos	S. André Capuava - S. Caetano

		Carapicuíba		Capão Redondo - Interlagos
2018	N. Senhora do Ó - Santana		1 fator	N. Senhora do Ó - Santana
	Mooca - Parque D. Pedro II	86,9%	Ibirapuera	Mooca - Parque D. Pedro II
	Ibirapuera - Parque D. Pedro II	Ibirapuera S. Caetano Parque D. Pedro II	Parque D. Pedro II S. Caetano Capão Redondo Interlagos Santana	S. André Capuava - S. Caetano
	Capão Redondo - Interlagos			Capão Redondo - Interlagos
2018qu	Cid. Universitária-USP-Ipen – Ibirapuera			
	N. Senhora do Ó - Santana	87,5%		
	Ibirapuera - Parque D. Pedro II	Ibirapuera S. Caetano Parque D. Pedro II	1 fator	Cid. Universitária-USP-Ipen - Ibirapuera
	S. André Capuava - S. Caetano	Parque D. Pedro II Cid.	Ibirapuera	
	Interlagos - Santo Amaro	Universitária-USP-Ipen Santana	Parque D. Pedro II S. Caetano	N. Senhora do Ó - Santana
	Mauá - S. André Capuava	EM-50 Sesc Itaquera S. André	EM-50 Sesc Itaquera S. André	Interlagos - Santo Amaro
	Ibirapuera - Santo Amaro	Capuava Santo Amaro N. Senhora do Ó	Capuava Santo Amaro	S. André Capuava - S. Caetano
	Capão Redondo - Santo Amaro	Ó Interlagos Capão Redondo		
	Parque D. Pedro II – Santana			
	Carapicuíba - Cid. Universitária-USP-Ipen			
2018fs	Mooca - Parque D. Pedro II	88,0%	1 fator	Mooca - Parque D. Pedro II
		Ibirapuera Cid.	Cid. Universitária-	Cid.

Cid. Universitária- USP-Ipen - Ibirapuera	Universitária- USP-Ipen Parque D. Pedro II	USP-Ipen Ibirapuera Parque D. Pedro II	Universitária- USP-Ipen - Ibirapuera
N. Senhora do Ó - Santana	S. Caetano Capão Redondo Interlagos	S. Caetano Capão Redondo Diadema	N. Senhora do Ó – Santana
Capão Redondo - Interlagos		Guarulhos-Paço Municipal Interlagos	Capão Redondo - Interlagos
Ibirapuera - Mooca		Itaim Paulista Mooca S. André	
Guarulhos- Pimentas - Itaim Paulista		Capuava Santana	
Ibirapuera - Parque D. Pedro II			
Guarulhos-Paço Municipal - Guarulhos- Pimentas			
Mauá - S. André Capuava			
Diadema - S. Caetano			

Mooca - Parque D.
Pedro II

2019

Ibirapuera - Parque D. Pedro II	86,6%	1 fator	Mooca - Parque D. Pedro II
Cid. Universitária- USP-Ipen - Ibirapuera	Ibirapuera S. Caetano Parque D. Pedro II	Ibirapuera Interlagos Parque D. Pedro II	Diadema – Interlagos
N. Senhora do Ó - Santana	Mooca Interlagos Diadema Cid.	S. Caetano Cid. Universitária- USP-Ipen	Cid. Universitária- USP-Ipen - Ibirapuera
Ibirapuera - Mooca	Universitária- USP-Ipen	Diadema Mooca Santana	N. Senhora do Ó - Santana
Diadema - Interlagos			

	Diadema - S. Caetano			
	Mooca - Parque D. Pedro II			
	Ibirapuera - Parque D. Pedro II			
	N. Senhora do Ó - Santana	91,0%	1 fator	
	Ibirapuera - Mooca	Ibirapuera S. Caetano Parque D. Pedro II	Ibirapuera Mooca Parque D. Pedro II	Mooca - Parque D. Pedro II
	S. André Capuava - S. Caetano	Mooca Santana	S. Caetano Cid.	Ibirapuera - Parque D. Pedro II
2019qu	Cid. Universitária- USP-Ipen - Ibirapuera	Santo Amaro Interlagos S. André Capuava Cid.	Universitária- USP-Ipen Diadema Interlagos S. André Capuava	N. Senhora do Ó - Santana
	Diadema - S. Caetano	Universitária- USP-Ipen	S. Bernardo- Centro	S. André Capuava - S. Caetano
	Interlagos - Santo Amaro	Diadema	Santo Amaro	
	Parque D. Pedro II – Santana			
	Ibirapuera - Santana			
	Cid. Universitária- USP-Ipen - Ibirapuera	86,4%	1 fator	Cid. Universitária- USP-Ipen – Ibirapuera
	Mooca - Parque D. Pedro II	Ibirapuera S. Caetano	Capão Redondo Diadema	Mooca - Parque D. Pedro II
	N. Senhora do Ó - Santana	Cid. Universitária- USP-Ipen	Ibirapuera Interlagos	
2019fs	Capão Redondo - Interlagos	Parque D. Pedro II Interlagos	Cid. Universitária- USP-Ipen	N. Senhora do Ó - Santana
	Ibirapuera - Parque D. Pedro II	Capão Redondo Diadema	Mooca	Capão Redondo - Interlagos
	Diadema - S.			

	Caetano			
	Cid. Universitária- USP-Ipen-Parque D. Pedro II			
	Interlagos - S. Bernardo-Centro			
	S. André Capuava - S. Caetano			
	Mooca - Parque D. Pedro II			Mooca - Parque D. Pedro II
2020	Guarulhos- Pimentas - Itaim Paulista	86,4%	1 fator	Guarulhos- Pimentas - Itaim Paulista
	Diadema - Interlagos		Ibirapuera Mooca Parque D. Pedro II	Diadema - Interlagos
	N. Senhora do Ó - Santana		Mooca Diadema	N. Senhora do Ó - Santana
	Parque D. Pedro II - Santana		Parque D. Pedro II Diadema Interlagos S. Caetano	
	Mooca-Parque D. Pedro II			Mooca - Parque D. Pedro II
2020qu	Diadema- Interlagos	87,0%	1 fator	N. Senhora do Ó - Santana
	Mauá - S. André Capuava		Diadema Ibirapuera Interlagos S. Caetano Cid. Universitária- USP-Ipen	S. André Capuava - S. Caetano
	Capão Redondo - Interlagos		Mooca Parque D. Pedro II	Ibirapuera - Parque D. Pedro II
	Ibirapuera - S. Caetano			
2020fs	Mooca - Parque D. Pedro II	85,1%	1 fator	Mooca - Parque D. Pedro II
	Guarulhos - Pimentas-Itaim Paulista		Ibirapuera Parque D. Pedro II Diadema Interlagos	Cid. Universitária- USP-Ipen - Ibirapuera
	Diadema - Interlagos		Mooca S. Caetano	Diadema -

				Interlagos
				Guarulhos- Pimentas - Itaim Paulista
	N. Senhora do Ó - Santana		1 fator	N. Senhora do Ó - Santana
		86,0%		Ibirapuera - Parque D. Pedro II
2016-2020	Ibirapuera - Parque D. Pedro II	Ibirapuera Parque D. Pedro II	Ibirapuera Parque D. Pedro II	
	Mooca - Parque D. Pedro II	S. Caetano Interlagos	S. Caetano Diadema Interlagos	Diadema - Interlagos
	Diadema - Interlagos		Santana	Guarulhos- Pimentas - Itaim Paulista

Para todos os períodos considerados a variância explicada pelo primeiro componente principal é expressiva, superior a 85%. Isso significa que o primeiro componente representa uma quantidade significativa das informações contidas nas variáveis originais, representando a maior parte da variabilidade do sistema. Quando se comparam os pesos do primeiro componente principal, as estações que possuem maiores pesos para a maioria dos períodos estudados são Ibirapuera (V8), Parque D. Pedro (V15) e S. Caetano (V19). Estas são, portanto, as estações mais importantes do sistema, quando se refere à contribuição para a variância. A estação com menor peso no primeiro componente principal para todas as análises é V16 (Pico do Jaraguá). Pode-se interpretar que se trata de uma estação com medições mais isoladas do restante do sistema, não se correlacionando com as demais estações.

No caso da análise de fatores, para a maior parte dos períodos, as estações de monitoramento que apresentam a maior carga para o primeiro fator são Ibirapuera (V8) e Parque D. Pedro II (V15), que possuem escala de representatividade urbana e bairro, respectivamente. O primeiro fator para todas as análises representou a maior proporção da variância explicada. E, novamente, assim como evidenciado no PCA, a estação Pico do Jaraguá apresentou a menor carga no primeiro fator. Outras estações com cargas mais baixas são Itaquera e Guarulhos-Pimentas. Com estes resultados, não foi possível chegar em uma interpretação sobre o significado do primeiro fator.

Tanto as análises das matrizes de correlação dos dados quanto as análises de *clusters* para todos os períodos mostraram que alguns pares de estações se repetem com maior

frequência como os pares mais correlacionados, sendo eles: Mooca (V13) e Parque D. Pedro II (V15); N. Senhora do Ó (V14) e Santana (V20); Ibirapuera (V8) e Cid. Universitária-USP-Ipen (V3). Quando se comparam os períodos quentes e úmidos com os frios e secos, nota-se que os resultados obtidos nos períodos quentes e úmidos são mais difusos, ou seja, aparecem diferentes pares de estações correlacionadas ano a ano. Este comportamento é diferente nos períodos frios e secos, em que os mesmos pares praticamente se repetem todos os anos: Mooca (V13) e Parque D. Pedro II (V15); N. Senhora do Ó (V14) e Santana (V20); Ibirapuera (V8) e Cid. Universitária-USP-Ipen (V3); Capão Redondo (V1) e Interlagos (V9). A hipótese é que isso ocorre pois nos períodos frios e secos há maior estabilidade atmosférica, permitindo uma menor variação das concentrações de ozônio medidas pelas estações de monitoramento, variação esta afetada principalmente pelas mudanças das condições meteorológicas.

A Figura 30 mostra a quantidade de períodos em que os pares mais correlacionados se repetem nas análises de *clusters*, quando se elegem para cada análise os 4 pares mais correlacionados.

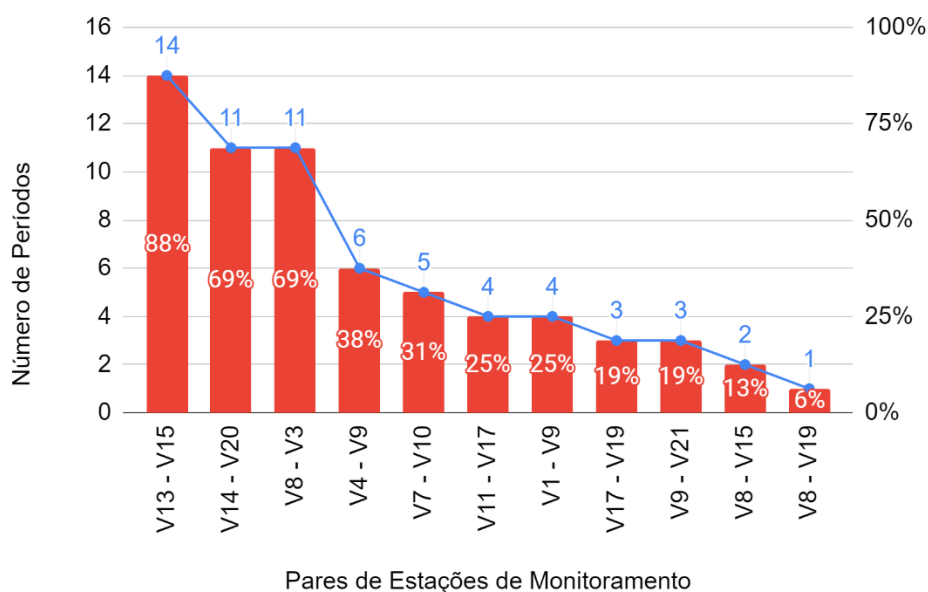


Figura 30 - Número de períodos com alto grau de similaridade para cada par de estações de monitoramento de O₃, obtido na Análise de *Clusters*

De uma maneira geral, apesar de haver 3 pares que se repetem como os mais correlacionados para todo o período analisado (Mooca-Parque D. Pedro II, N. Senhora do Ó - Santana e Ibirapuera-Cid. Universitária-USP-Ipen), existem vários pares com altas correlações para o ozônio, acima de 0,95. Acredita-se que isso se deve à grande influência da

radiação solar na formação deste poluente, fazendo com que os perfis diários de concentração sejam semelhantes entre as estações.

4.2. MP₁₀

Diferentemente do ozônio, o material particulado é um poluente primário, ou seja, emitido principalmente por fontes veiculares na RMSP. O período do ano que geralmente apresenta as maiores concentrações de MP₁₀ e MP_{2,5} é de maio a setembro, por ser um período seco e sujeito a eventos de calmaria e inversão térmica no estado de São Paulo (CETESB, 2021).

Entre a Figura 31 e a Figura 35 estão representadas as distribuições espaciais das concentrações médias anuais de MP₁₀ em cada período.

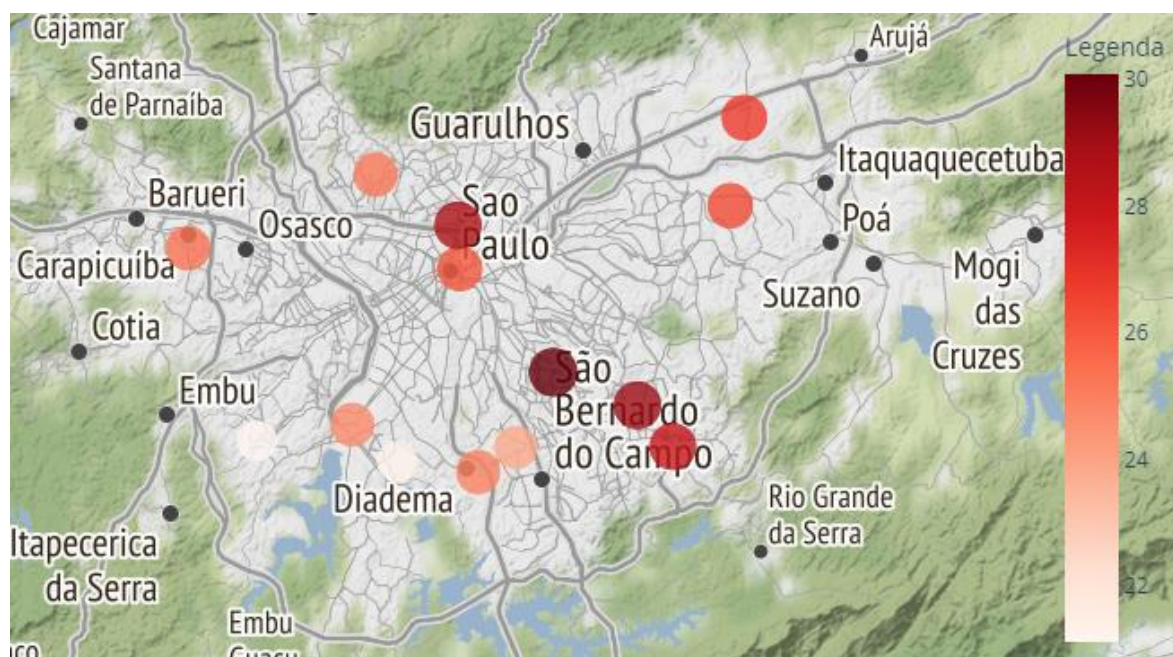


Figura 31 - Concentração média anual de MP₁₀, em µg/m³, em 2016.

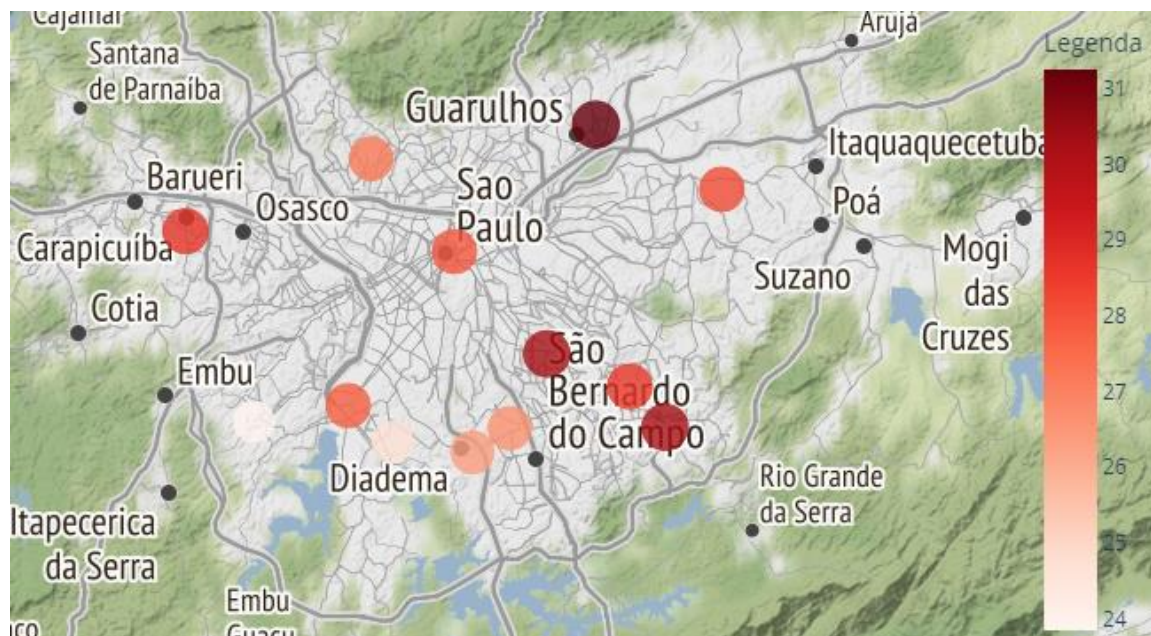


Figura 32 - Concentração média anual de MP₁₀, em µg/m³, em 2017.

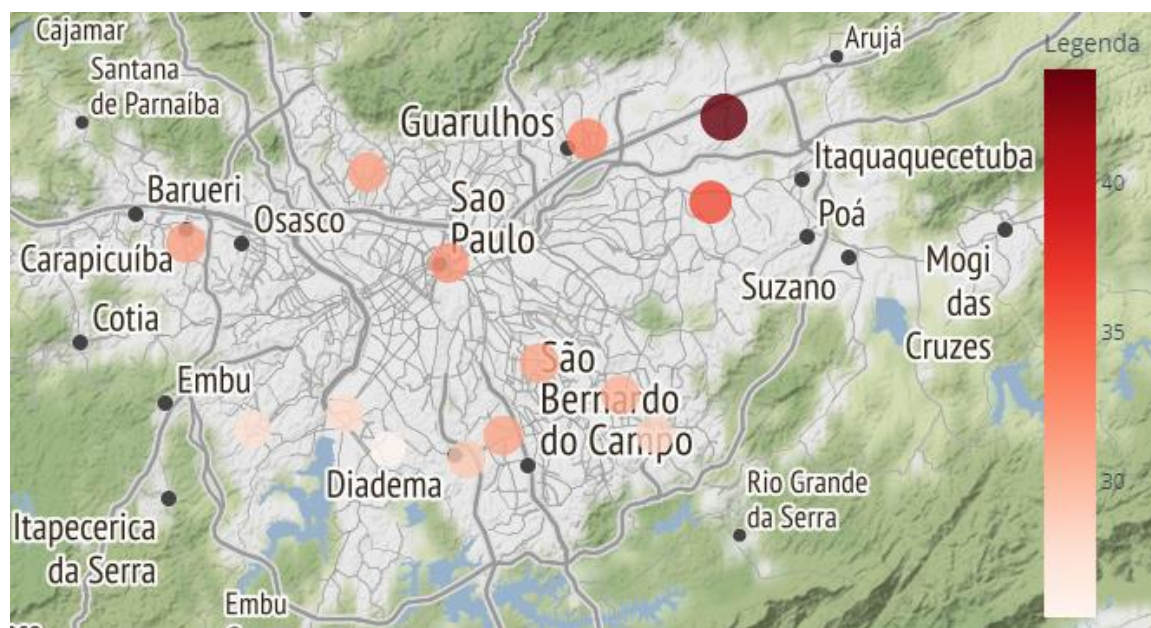


Figura 33 - Concentração média anual de MP₁₀, em µg/m³, em 2018.

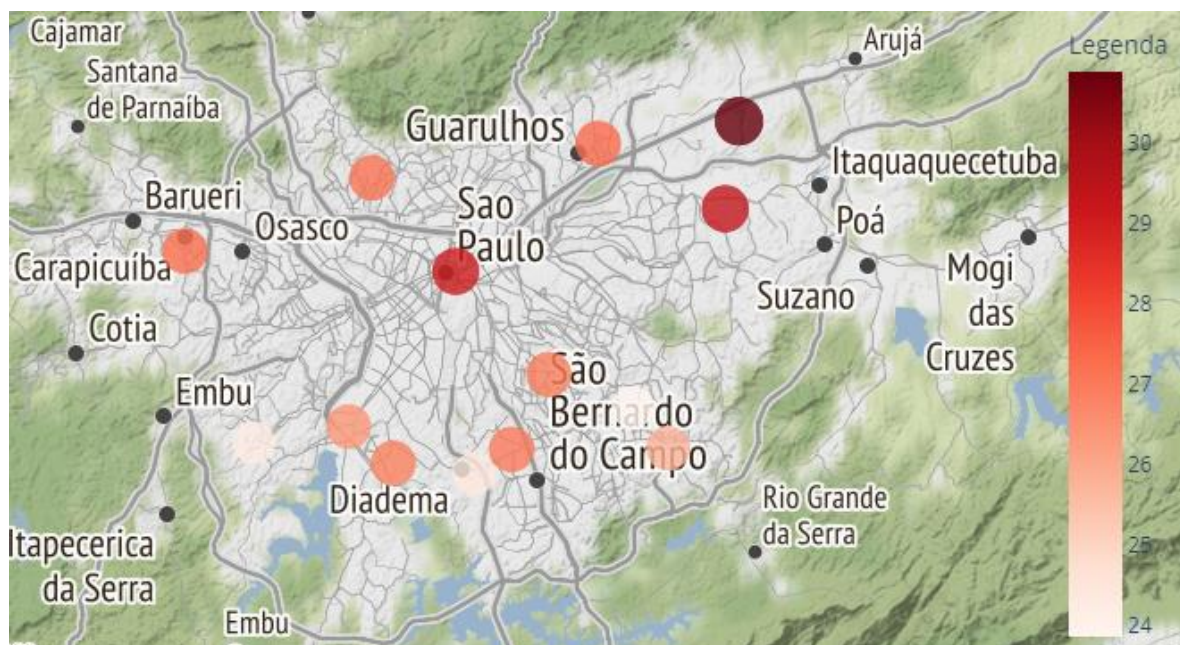


Figura 34 - Concentração média anual de MP₁₀, em µg/m³, em 2019.

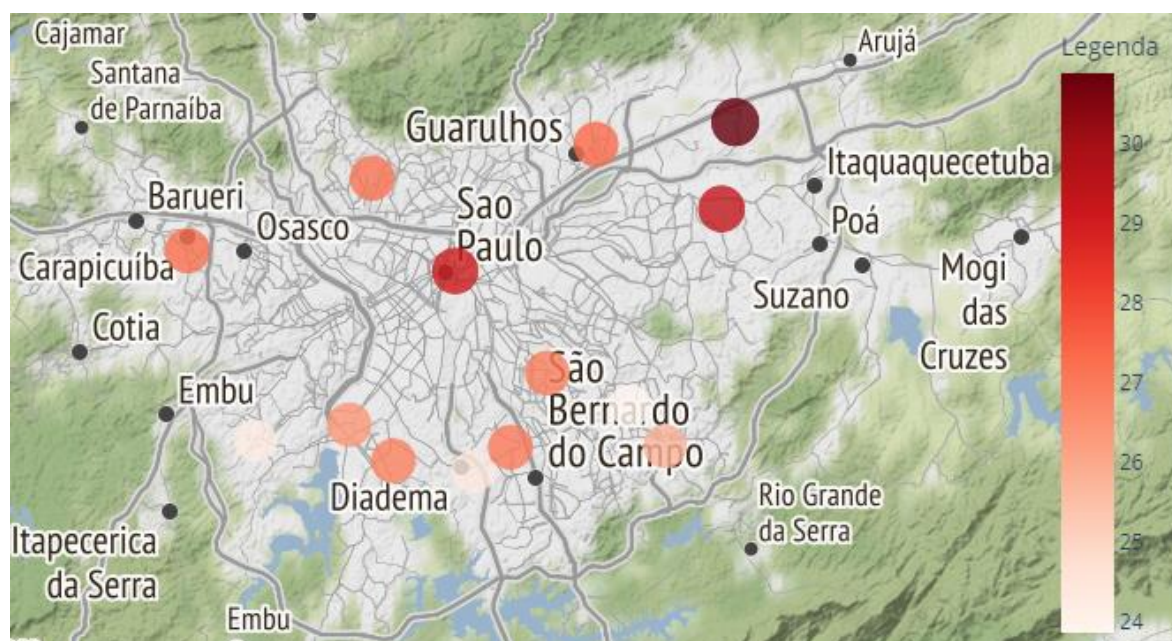


Figura 35 - Concentração média anual de MP₁₀, em µg/m³, em 2020.

Entre a Figura 36 e a Figura 40 estão representadas as distribuições espaciais das concentrações médias de MP₁₀ nos períodos quentes e úmidos.

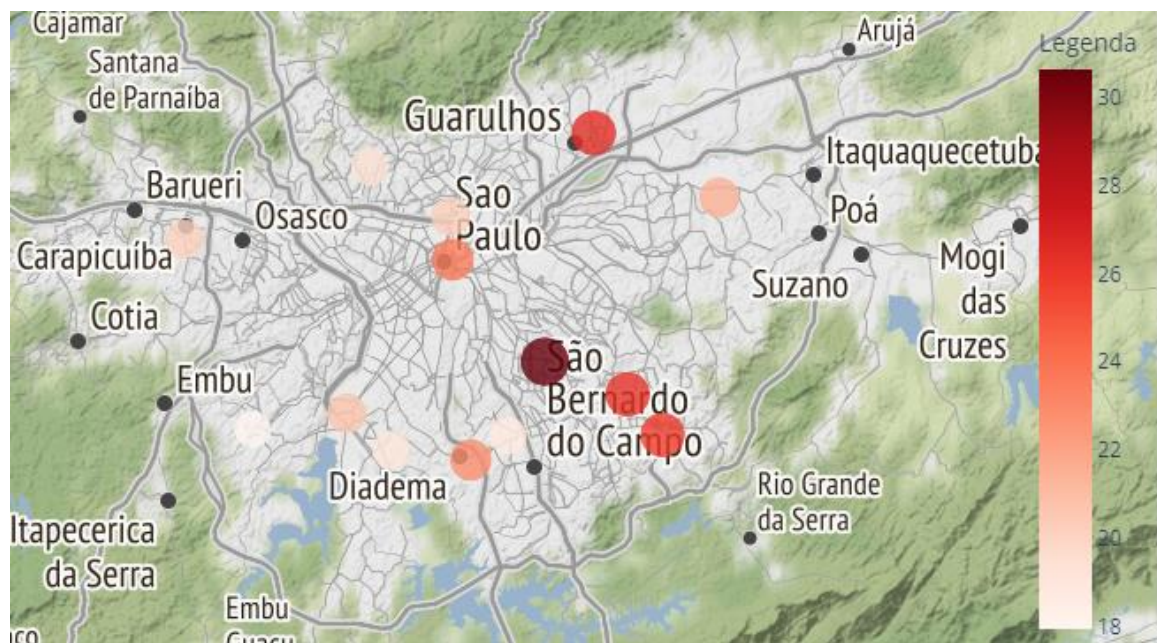


Figura 36 - Concentração média de MP_{10} , em $\mu g/m^3$, no período quente e úmido em 2016.

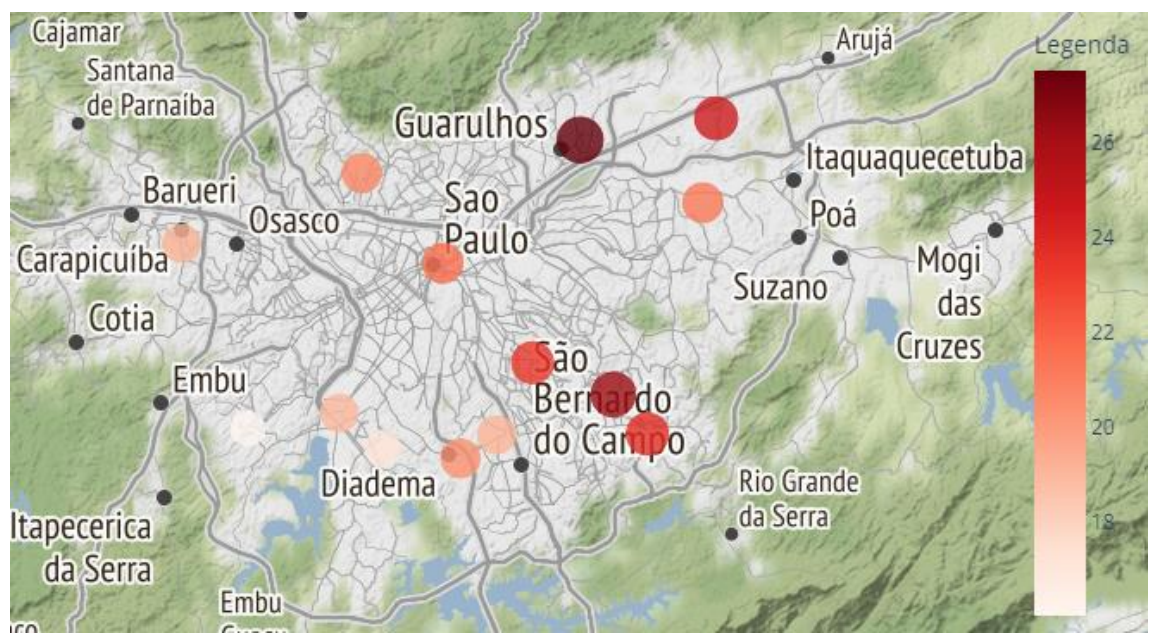


Figura 37 - Concentração média de MP_{10} , em $\mu g/m^3$, no período quente e úmido em 2017.

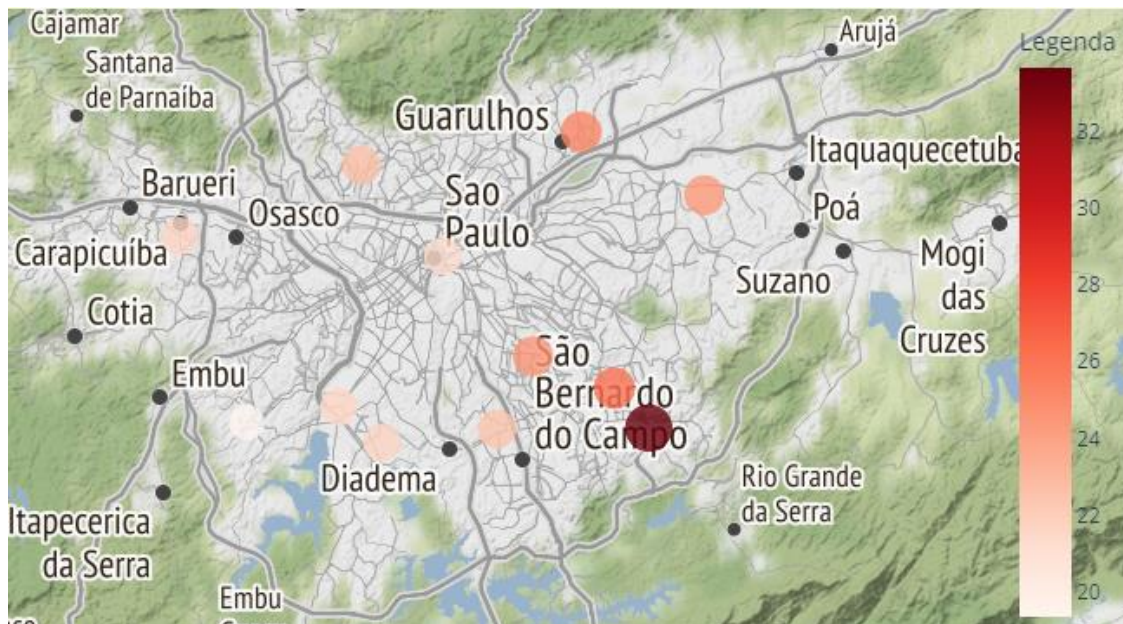


Figura 38 - Concentração média de MP10, em µg/m³, no período quente e úmido em 2018.

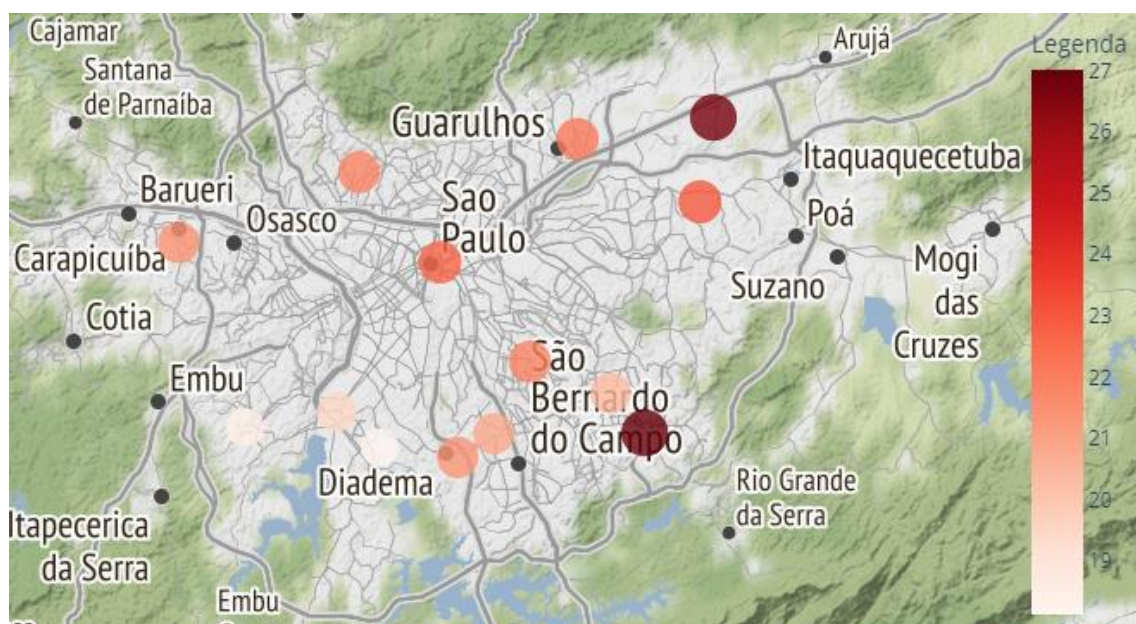


Figura 39 - Concentração média de MP10, em µg/m³, no período quente e úmido em 2019.

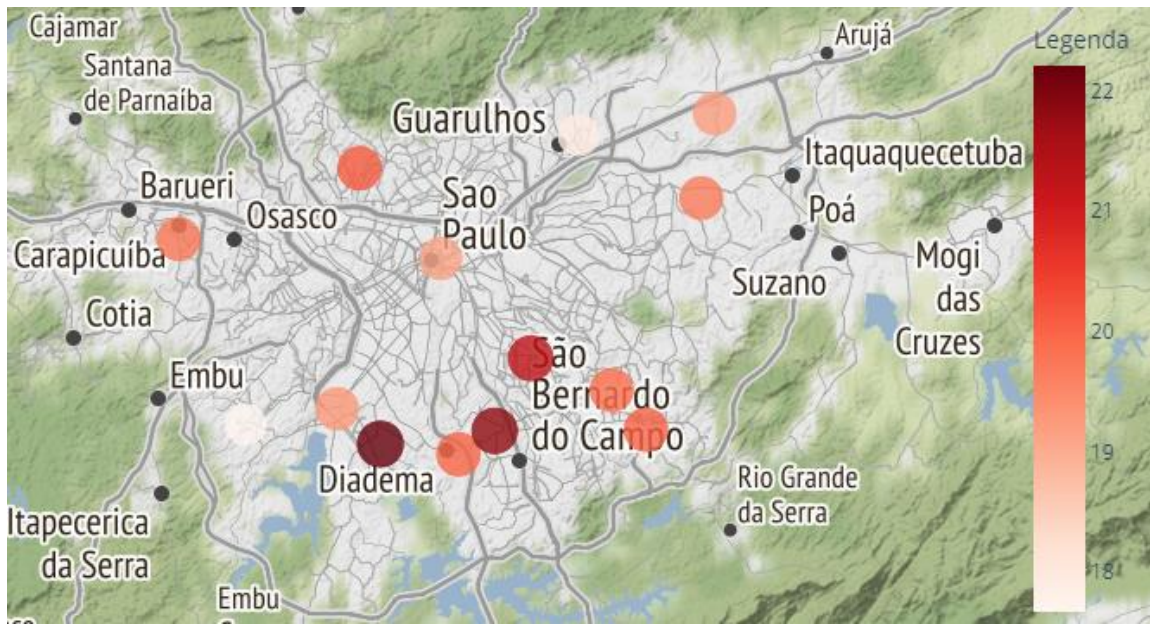


Figura 40 - Concentração média de MP₁₀, em µg/m³, no período quente e úmido em 2020.

Entre a Figura 41 e a Figura 45 estão representadas as distribuições espaciais das concentrações médias de MP₁₀ nos períodos frios e secos.

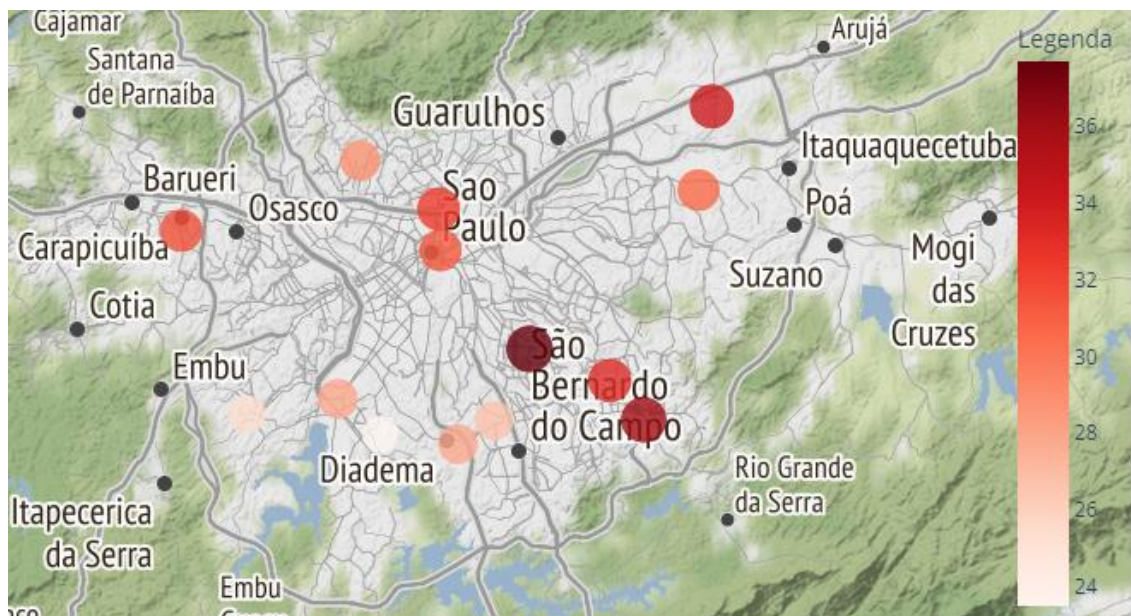


Figura 41 - Concentração média de MP₁₀, em µg/m³, no período frio e seco em 2016.

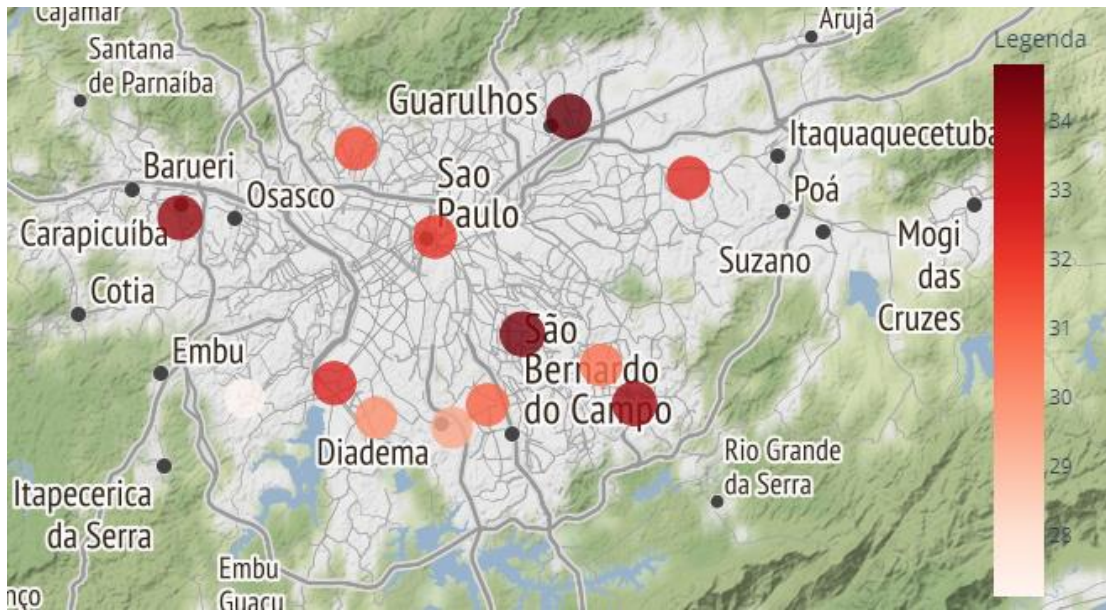


Figura 42 - Concentração média de MP_{10} , em $\mu g/m^3$, no período frio e seco em 2017.

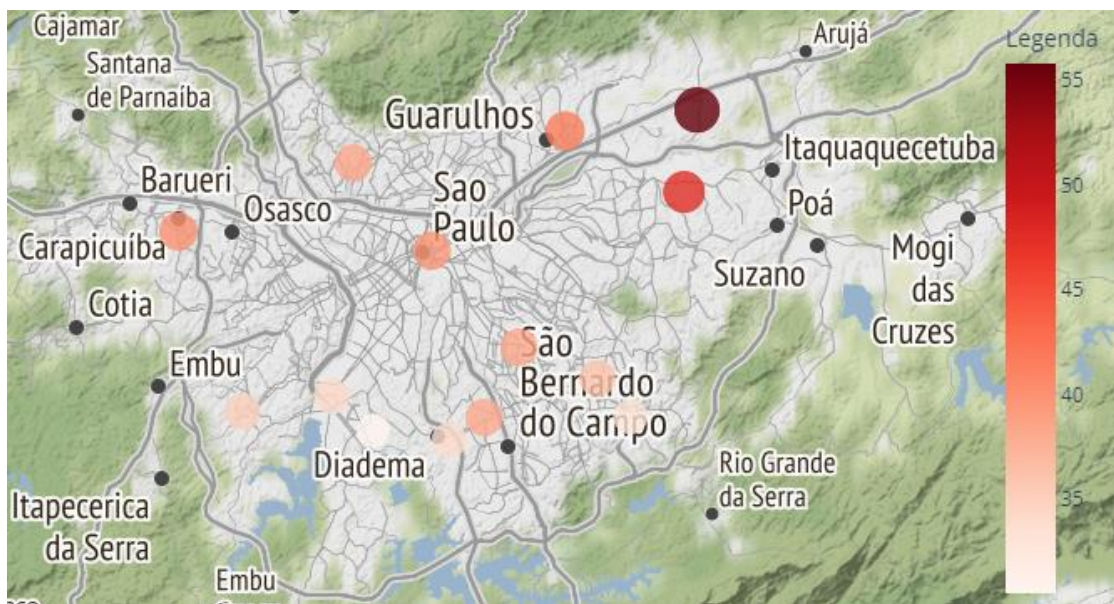


Figura 43 - Concentração média de MP_{10} , em $\mu g/m^3$, no período frio e seco em 2018.

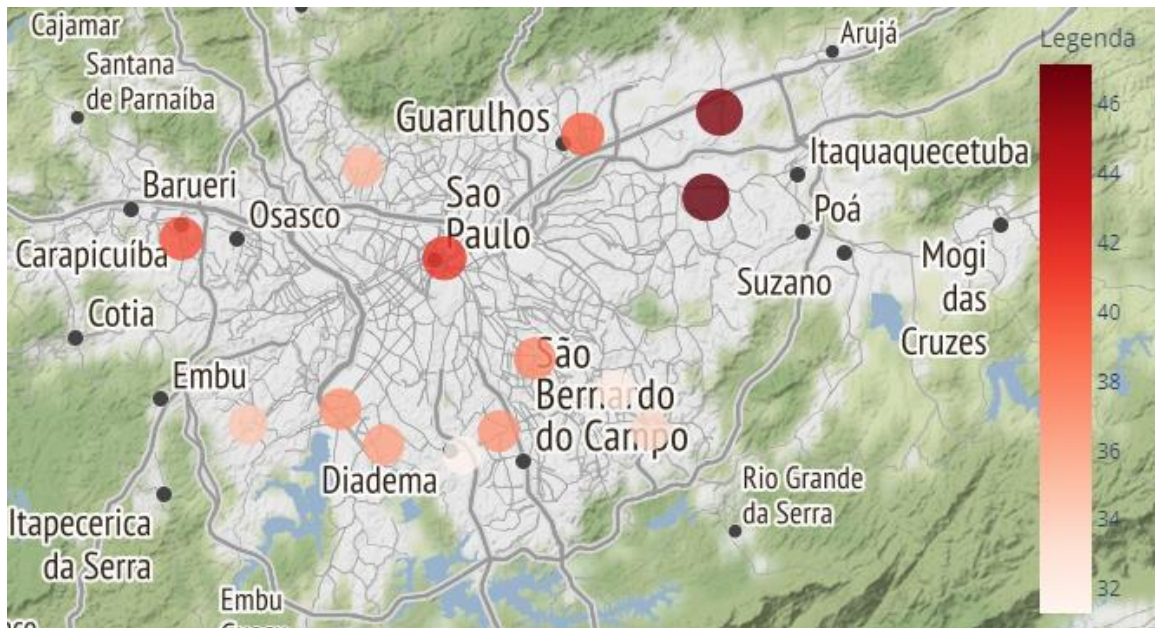


Figura 44 - Concentração média de MP_{10} , em $\mu\text{g}/\text{m}^3$, no período frio e seco em 2019.

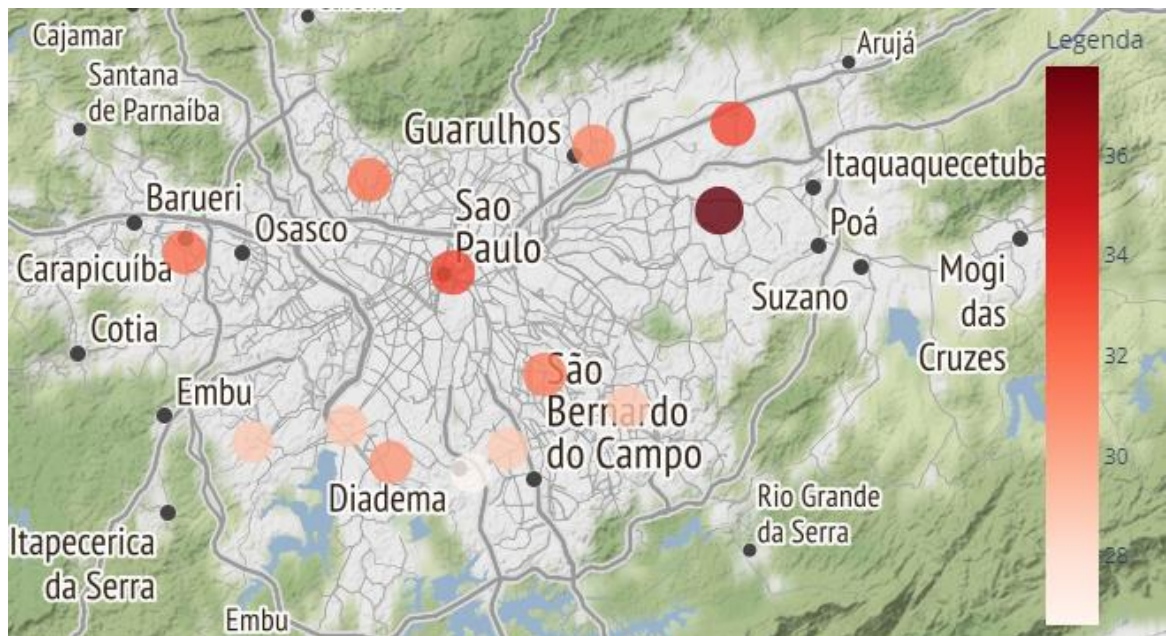


Figura 45 - Concentração média de MP_{10} , em $\mu\text{g}/\text{m}^3$, no período frio e seco em 2020.

No caso das concentrações médias de MP_{10} , a estação Guarulhos-Pimentas destaca-se como uma das estações com os maiores valores. Ela tem escala de representatividade de bairro.

Nota-se que nos períodos frios e secos as médias são superiores aos períodos quentes e úmidos, chegando à média máxima de $55 \mu\text{g}/\text{m}^3$ em 2018 e $46 \mu\text{g}/\text{m}^3$ em 2019. Pode-se

relacionar isso tanto às influências dos efeitos meteorológicos quanto ao transporte de cinzas oriundas de queimadas.

A Tabela 9 a seguir apresenta os resultados das análises multivariadas ano a ano. A coluna MC mostra os pares de estações mais correlacionados na matriz de correlação, com coeficiente acima de 0,80. No caso do PCA, a porcentagem indica a variância explicada pelo primeiro componente principal (PC1) e as estações com pesos no PC1 superiores a 0,28. A coluna AF mostra quantos fatores foram mantidos para análise e as estações com maiores *loadings* no primeiro fator, acima de 0,80, que explica a maior proporção de variância dos dados. E, por fim, a coluna AC contém os pares de estações com maior grau de similaridade.

Tabela 9 - Resumo dos principais resultados das técnicas de análise estatística multivariada para o poluente MP₁₀.

Base de Dados	MC	PCA	AF	AC
2016	Capão Redondo - Santo Amaro	56,8%	1 fator	Capão Redondo - Santo Amaro
		Parque D. Pedro II Santo Amaro S. Caetano Interlagos Capão Redondo	Capão Redondo Interlagos Parque D. Pedro II S. Caetano Santo Amaro	Parque D. Pedro II - S. Caetano
2016qu	Sem correlações acima de 0,80.	41,6%	3 fatores	Itaim Paulista - Parque D. Pedro II
		Santo Amaro S. Caetano S. Bernardo-Pauliceia Santana Diadema Mauá S. André Capuava	Sem loadings acima de 0,80.	S. Bernardo-Pauliceia - S. Caetano Diadema - S. André Capuava
2016fs	Capão Redondo - Santo Amaro	63,4%	1 fator	Capão Redondo - Santo Amaro
		Parque D. Pedro II Interlagos Santo Amaro S. Caetano Capão Redondo	Capão Redondo Diadema Interlagos Parque D. Pedro II S. Caetano Santo Amaro	Guarulhos-Pimentas - Itaim Paulista Interlagos-Parque D. Pedro II
2017	N. Senhora do Ó - Parque D. Pedro II	62,2%	1 fator	N. Senhora do Ó - Parque D. Pedro II
		Parque D. Pedro	N. Senhora do	

	Parque D. Pedro II - Santo Amaro	II N. Senhora do Ó Santo Amaro S. Caetano S. Bernardo-Pauliceia Capão Redondo	Ó Parque D. Pedro II S. Bernardo-Pauliceia S. Caetano Santo Amaro	Capão Redondo - Santo Amaro S. Bernardo-Pauliceia - S. Caetano
2017qu	Sem correlações acima de 0,80.	48,6%	2 fatores	N. Senhora do Ó - Parque D. Pedro II Capão Redondo - Santo Amaro
		Parque D. Pedro II Santo Amaro S. Caetano S. Bernardo-Pauliceia N. Senhora do Ó Guarulhos-Pimentas	Parque D. Pedro II	S. Bernardo-Pauliceia - S. Caetano Diadema - S. André Capuava
2017fs	N. Senhora do Ó - Parque D. Pedro II Parque D. Pedro II - Santo Amaro	63,5%	1 fator	N. Senhora do Ó-Parque D. Pedro II
		Parque D. Pedro II N. Senhora do Ó S. Caetano Santo Amaro S. Bernardo-Pauliceia	N. Senhora do Ó Parque D. Pedro II S. Bernardo-Pauliceia S. Caetano Santo Amaro	Capão Redondo-Santo Amaro S. Bernardo-Pauliceia - S. Caetano
2018	Capão Redondo - Santo Amaro	70,9%	1 fator	Capão Redondo - Santo Amaro
	Capão Redondo - Carapicuíba		Capão Redondo Carapicuíba Guarulhos-Paco Municipal Interlagos	Guarulhos-Paco Municipal - Itaim Paulista
	Interlagos - Santo Amaro	Parque D. Pedro II Itaim Paulista S. Caetano	Itaim Paulista N. Senhora do Ó Parque D. Pedro II S. Bernardo-Pauliceia S. Caetano Santo Amaro	Parque D. Pedro II - S. Caetano
2018qu	Sem correlações	53,2%	2 fatores	S. Caetano -

	acima de 0,80.	Capão Redondo Santo Amaro Parque D. Pedro II S. Caetano Itaim Paulista S. Bernardo- Pauliceia N. Senhora do Ó	Capão Redondo	Santo Amaro Capão Redondo - Itaim Paulista
			1 fator	
2018fs	Capão Redondo - Santo Amaro		Capão Redondo Carapicuíba Diadema Guarulhos-Paco Municipal Interlagos Itaim Paulista N. Senhora do Ó Parque D. Pedro II S. Bernardo- Pauliceia S. Caetano Santo Amaro	Capão Redondo - Santo Amaro Guarulhos-Paco Municipal - Itaim Paulista Parque D. Pedro II - S. Caetano
	Interlagos - Santo Amaro	72,1%		
	Capão Redondo - Carapicuíba			
	Parque D. Pedro II - S. Caetano			
			1 fator	
2019	Capão Redondo - Santo Amaro		Capão Redondo Interlagos Itaim Paulista Mauá Parque D. Pedro II S. André Capuava S. Bernardo- Pauliceia S. Caetano Santo Amaro	Capão Redondo - Santo Amaro Parque D. Pedro II - S. Caetano Guarulhos-Paco Municipal - Santo Amaro
	Parque D. Pedro II - S. Caetano	69,8%		
	Interlagos - Santo Amaro			
	S. Caetano - Santo Amaro			
			3 fatores	
2019qu	Sem correlações acima de 0,80.	Parque D. Pedro II S. Caetano Itaim Paulista Capão Redondo	Itaim Paulista Parque D. Pedro II S. Caetano	Guarulhos-Paco Municipal - Guarulhos-Pimentas Capão Redondo - Santo Amaro

N. Senhora do
Ó

N. Senhora do
Ó - Parque D.
Pedro II

Mauá - S. André
Capuava

2019fs	Parque D. Pedro II - S. Caetano			
	Parque D. Pedro II - Santo Amaro			
	S. Bernardo- Pauliceia - S. Caetano		1 fator	
	Interlagos-Santo Amaro		Capão Redondo Carapicuíba Guarulhos-Paco Municipal	Parque D. Pedro II - S. Caetano
	Itaim Paulista - S. André Capuava	77,4%	Interlagos Itaim Paulista Mauá	Interlagos - Santo Amaro
	Parque D. Pedro II - S. Bernardo- Pauliceia	Parque D. Pedro II	N. Senhora do Ó Parque D. Pedro II	Itaim Paulista - S. André Capuava
	Capão Redondo - Santo Amaro		S. André Capuava	
	Mauá - S. Caetano		S. Bernardo- Pauliceia	
	S. Bernardo- Pauliceia - Santo Amaro		S. Caetano Santo Amaro	
	S. Caetano - Santo Amaro			
2020	Capão Redondo - Santo Amaro	69,8%	1 fator	Capão Redondo - Santo Amaro
	Parque D. Pedro II - S. Caetano	S. Caetano Santo Amaro	Capão Redondo Guarulhos-Paco Municipal	Parque D. Pedro II - S. Caetano
	Interlagos - Santo Amaro	Capão Redondo Parque D. Pedro II	Interlagos Itaim Paulista	Guarulhos-Paco Municipal - Itaim Paulista
	S. Caetano - Santo Amaro	N. Senhora do Ó Guarulhos-Paco Municipal	Mauá Parque D. Pedro II S. André	

			Capuava S. Bernardo- Pauliceia S. Caetano Santo Amaro	
2020qu	Sem correlações acima de 0,80.	50,1%		Capão Redondo - Santo Amaro
		Mauá Santo Amaro S. Caetano Guarulhos-Paco Municipal Capão Redondo Itaim Paulista	2 fatores Mauá S. Caetano Santo Amaro	Mauá - S. Caetano Guarulhos-Paco Municipal - Itaim Paulista
2020fs	Interlagos - Santo Amaro			
	Capão Redondo - Santo Amaro		1 fator	
	Parque D. Pedro II - Santo Amaro	72,1%	Capão Redondo Carapicuíba Guarulhos-Paco Municipal Interlagos Itaim Paulista N. Senhora do Ó	Interlagos - Santo Amaro Parque D. Pedro II - S. Caetano Capão Redondo - Carapicuíba
	Parque D. Pedro II - S. Caetano	Santo Amaro Parque D. Pedro II	Parque D. Pedro II	Guarulhos-Paco Municipal - Guarulhos- Pimentas
	S. Caetano - Santo Amaro	S. Caetano Interlagos Capão Redondo Itaim Paulista	S. Bernardo- Pauliceia S. Caetano Santo Amaro	
	Capão Redondo - Carapicuíba			
	Carapicuíba - Santo Amaro			
Capão Redondo - Interlagos		1 fator		
2016-2020	Capão Redondo - Santo Amaro	68,5%	Capão Redondo Carapicuíba Guarulhos-Paco Municipal Interlagos Itaim Paulista N. Senhora do Ó Parque D. Pedro II	Capão Redondo - Santo Amaro Parque D. Pedro II - S. Caetano Guarulhos-Paco Municipal - Itaim Paulista
		S. Caetano Parque D. Pedro II Santo Amaro Itaim Paulista Capão Redondo		

Para a maioria dos períodos considerados, a variância explicada pelo primeiro componente principal é superior a 65%, com exceção dos períodos com maior precipitação (quentes e úmidos). Isso pode ser explicado pela condição meteorológica favorável a menores níveis de particulado na atmosfera, prejudicando a correlação entre as medições das estações. Para os materiais particulados, diferentemente do ozônio, o período do ano de maior interesse para a análise é o de meses mais frios e secos.

As estações com maiores pesos no primeiro componente principal na PCA, na maior parte dos períodos, são S. Caetano (V16), Parque D. Pedro II (V13) e Santo Amaro (V18). Ou seja, no caso do MP₁₀, são essas as estações que mais contribuem para a variância do sistema.

No caso da análise de fatores, ao analisar as estações de monitoramento com maior carga para o primeiro fator, pode-se destacar para a maior parte dos períodos as estações Parque D. Pedro II (V13), S. Caetano (V16) e Santo Amaro (V18), que possuem escala de representatividade bairro, média e média, respectivamente. Já as estações com menores cargas no primeiro fator para a maioria dos períodos são Guarulhos-Pimentas (V5), N. Senhora do Ó (V12) e Itaim Paulista (V8). Dessa forma, assim como ocorreu para o poluente O₃, uma possível interpretação do primeiro fator é que ele está associado à proximidade de vias movimentadas.

Algo interessante que ocorreu na análise de fatores do MP₁₀ é que em todos os períodos quentes e úmidos houve mais de um fator com autovalor superior a 1. No caso dos períodos “2016qu” e “2019qu”, a análise resultou em 3 fatores, e no caso dos períodos de 2017, 2018 e 2020, surgiram 2 fatores.

Conforme demonstrado pelas tabelas do APÊNDICE F, não existe um padrão claro entre as cargas fatoriais do segundo e terceiro fator. Dessa forma, não foi possível chegar a conclusões satisfatórias a respeito da interpretação dos demais fatores além do primeiro, mas acredita-se que estes períodos tenham resultado em mais fatores devido à presença de chuvas que afetam a concentração do material particulado.

As análises das matrizes de correlação e as análises de *clusters* para todos os períodos mostram que o par de estações mais correlacionadas, que se repetem com maior frequência, é Capão Redondo (V1) e Santo Amaro (V18).

A Figura 46 mostra a quantidade de períodos em que os pares mais correlacionados se repetem nas análises de *clusters*, quando se elege para cada análise os pares mais correlacionados.

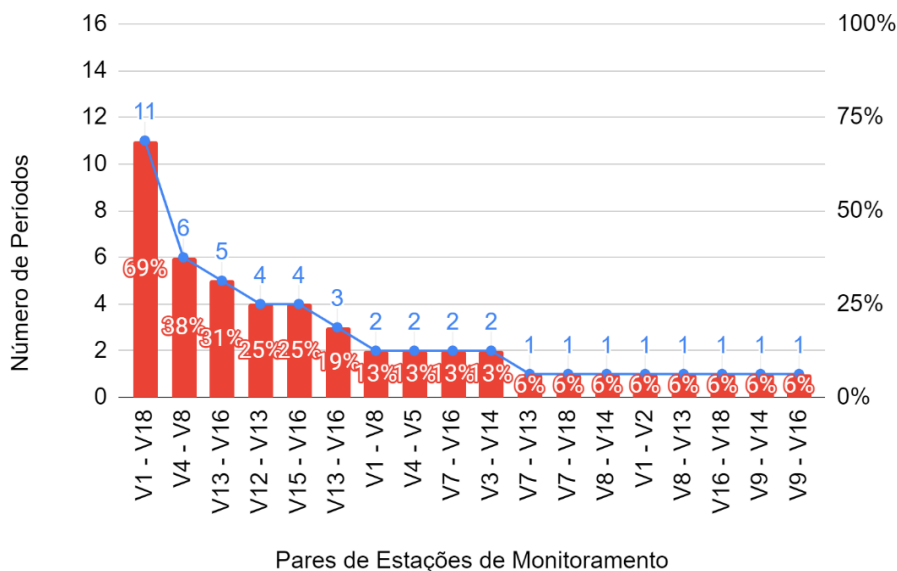


Figura 46 - Número de períodos para cada par de estações de monitoramento de MP₁₀ com alto grau de similaridade na Análise de Clusters

Pode-se concluir que o par de estações que estão correlacionadas de forma mais consistente ao longo do período de 2016 a 2020 é formado por Capão Redondo (V1) e Santo Amaro (V18). As medições da estação Santo Amaro representam boa parte da variabilidade do sistema como um todo, conforme confirmado pelas análises de componentes principais e análise de fatores.

4.3. MP_{2,5}

Dentre os poluentes analisados, o MP_{2,5} é o que possui o menor número de estações monitoradas, totalizando 12. Assim como para o poluente MP₁₀, o período do ano de maior interesse para a análise de MP_{2,5} é composto por meses frios e secos, devido à maior concentração desses poluentes na atmosfera.

As Figura 47 a Figura 51 mostram as concentrações médias anuais deste poluente para cada período.

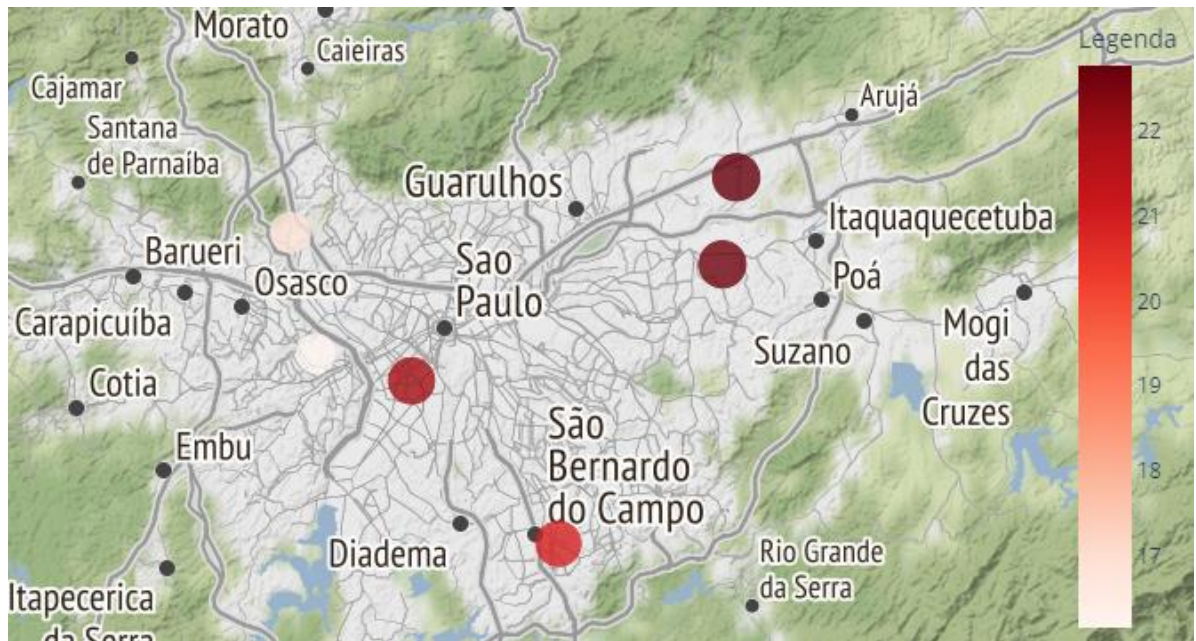


Figura 47 - Concentração média anual de MP_{2.5}, em µg/m³, em 2016.

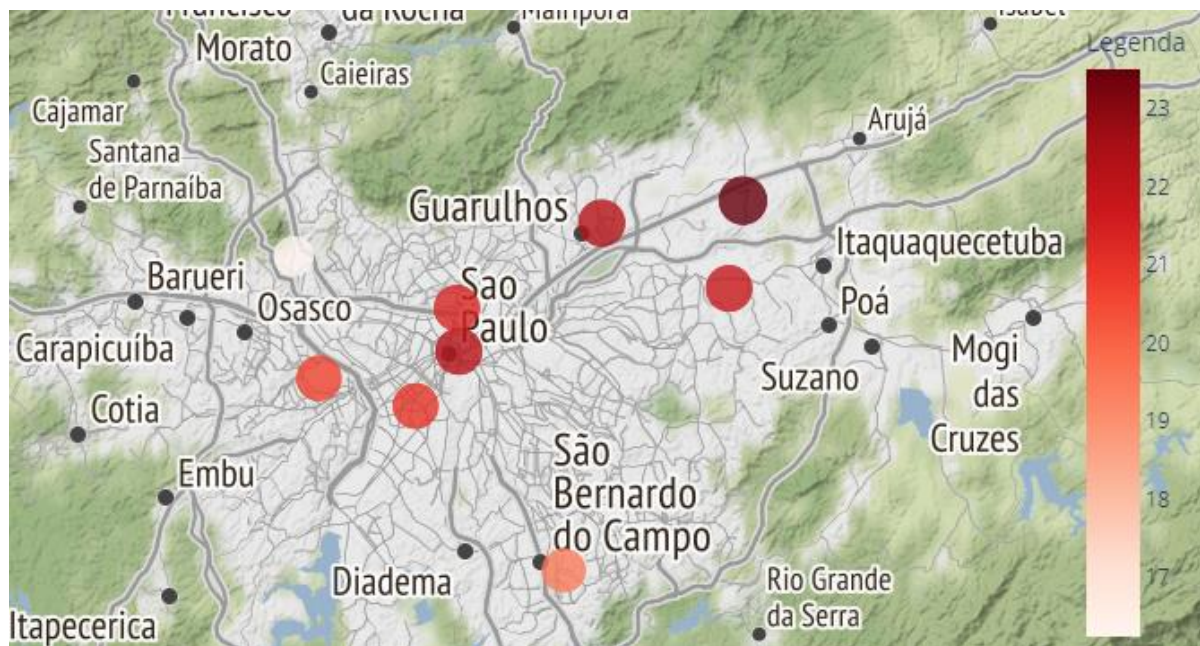


Figura 48 - Concentração média anual de MP_{2.5}, em µg/m³, em 2017.

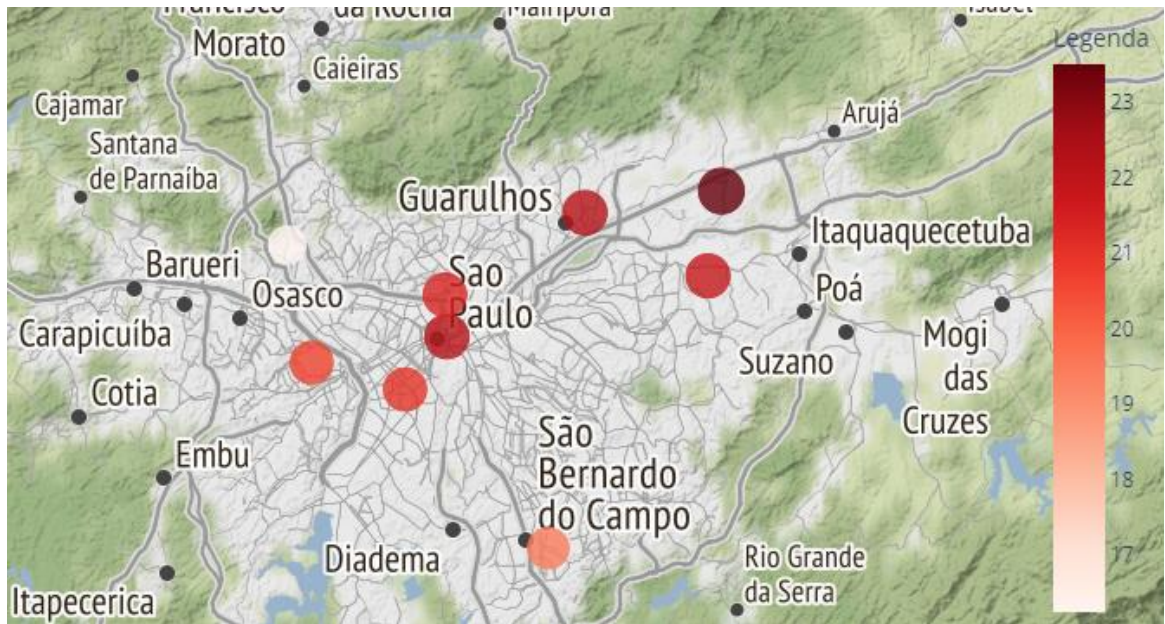


Figura 49 - Concentração média anual de $MP_{2.5}$, em $\mu g/m^3$, em 2018.

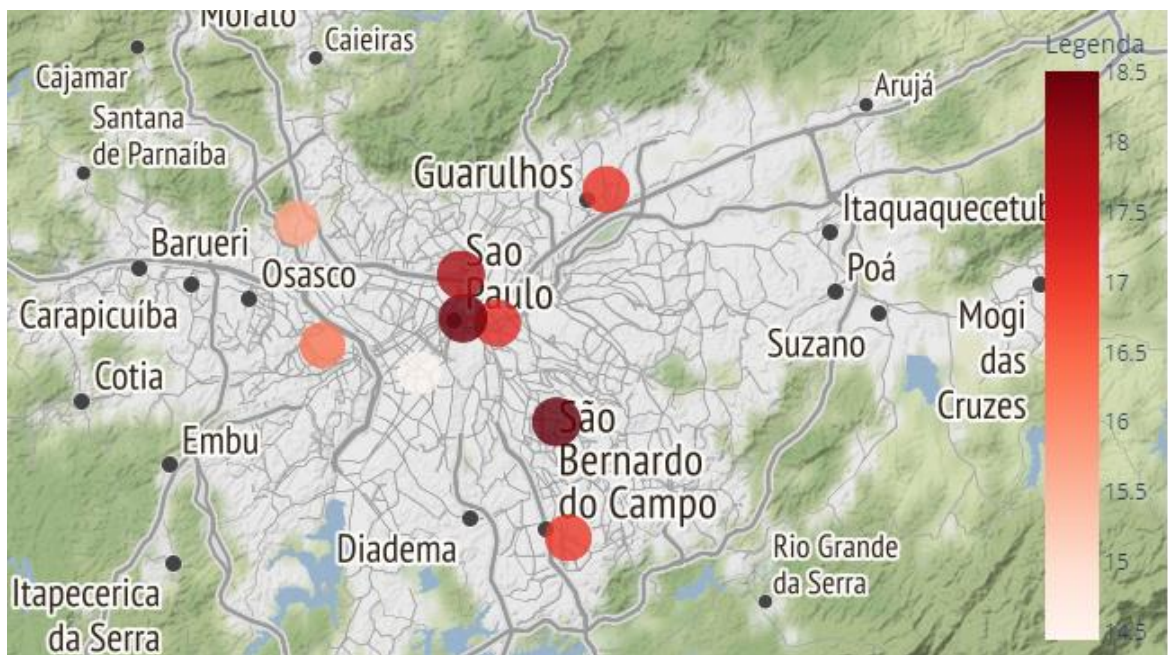


Figura 50 - Concentração média anual de $MP_{2.5}$, em $\mu g/m^3$, em 2019.

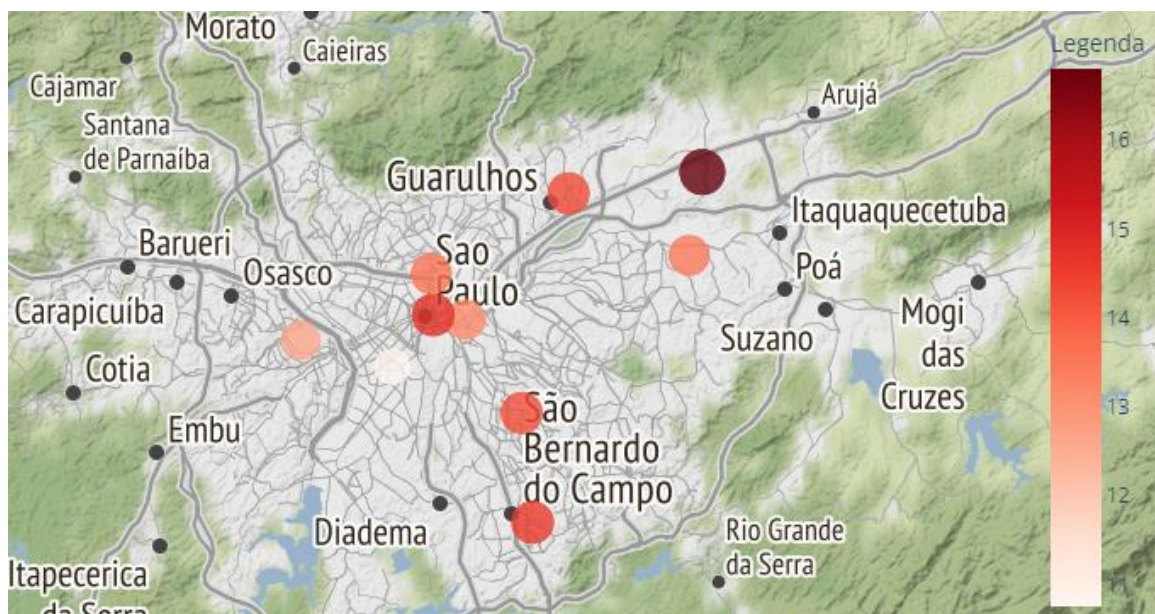


Figura 51 - Concentração média anual de MP2.5, em $\mu\text{g}/\text{m}^3$, em 2020.

Entre a Figura 52 e a Figura 56 estão representadas as distribuições espaciais das concentrações médias de MP_{2.5} nos períodos quentes e úmidos.

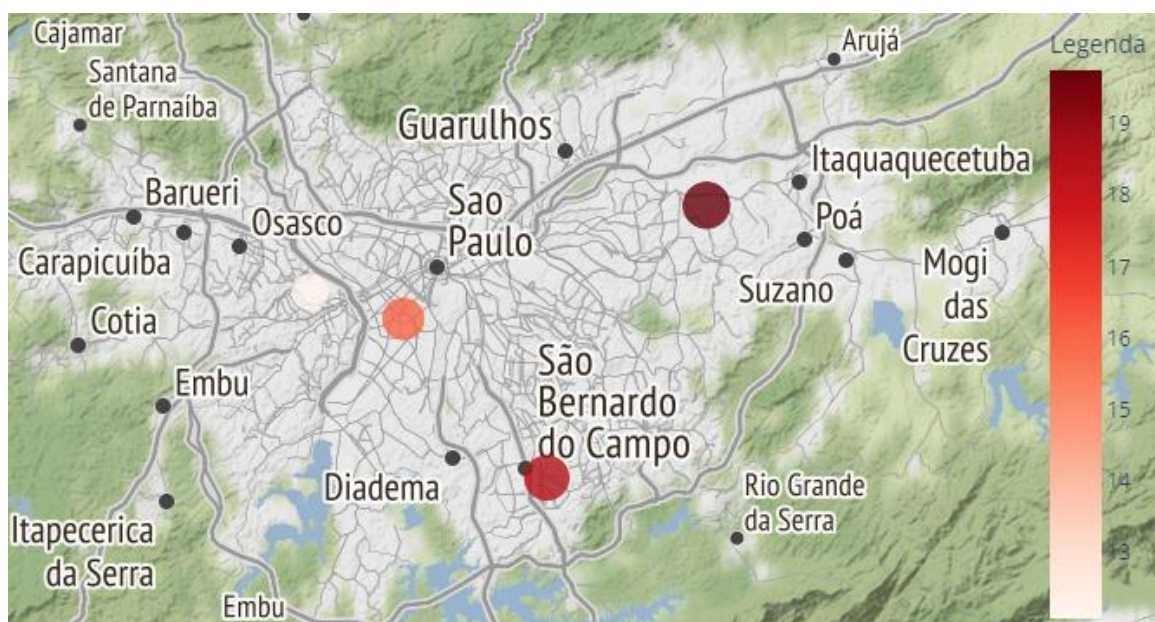


Figura 52 - Concentração média de MP_{2.5}, em $\mu\text{g}/\text{m}^3$, no período quente e úmido em 2016.

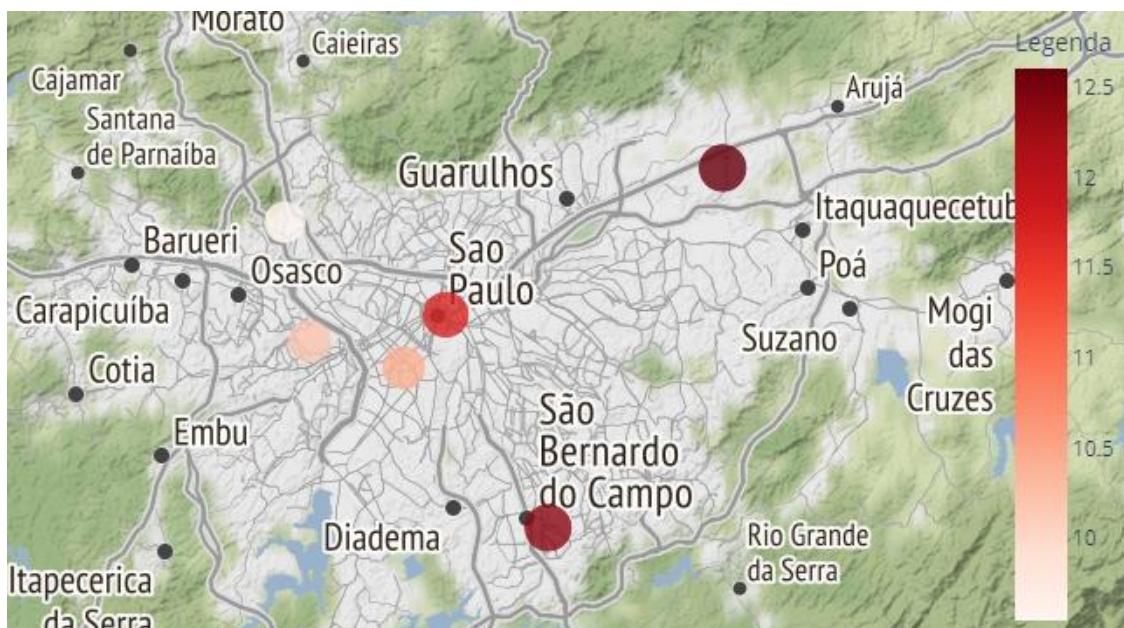


Figura 53 - Concentração média de MP_{2.5}, em µg/m³, no período quente e úmido em 2017.

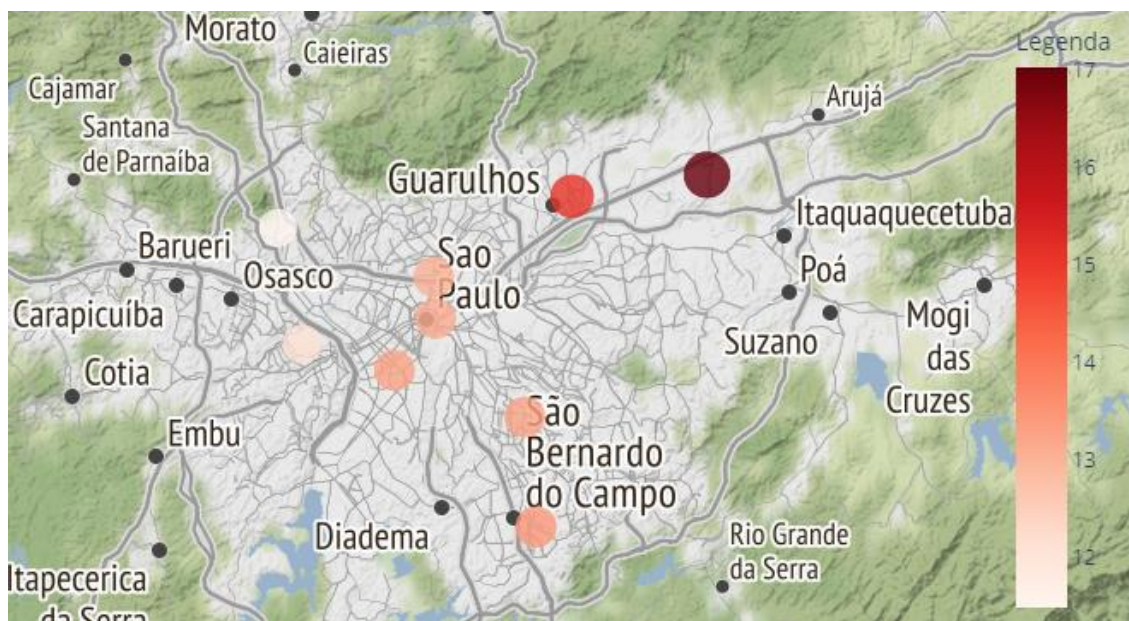


Figura 54 - Concentração média de MP_{2.5}, em µg/m³, no período quente e úmido em 2018.

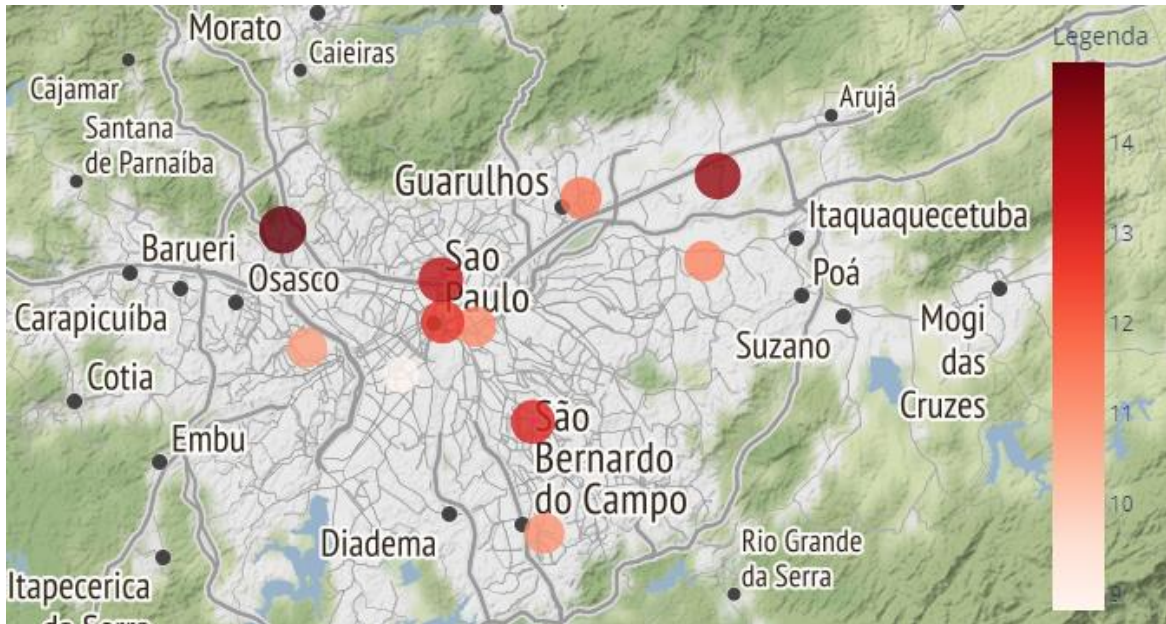


Figura 55 - Concentração média de MP_{2.5}, em µg/m³, no período quente e úmido em 2019.

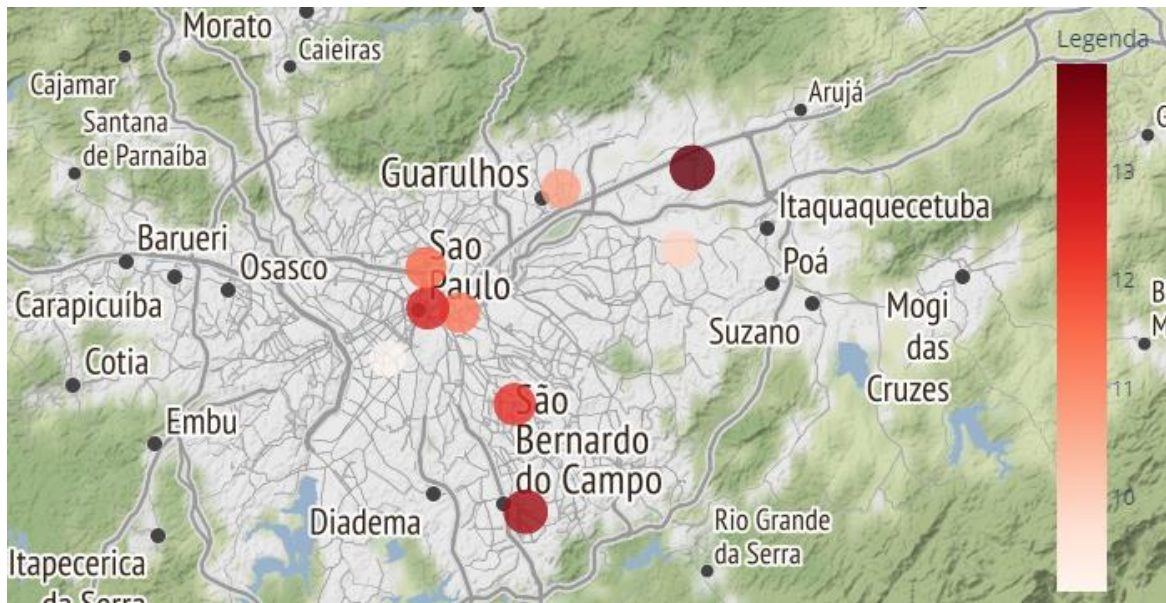


Figura 56 - Concentração média de MP_{2.5}, em µg/m³, no período quente e úmido em 2020.

Entre a Figura 57 e a Figura 61 estão representadas as distribuições espaciais das concentrações médias de MP_{2.5} nos períodos frios e secos.

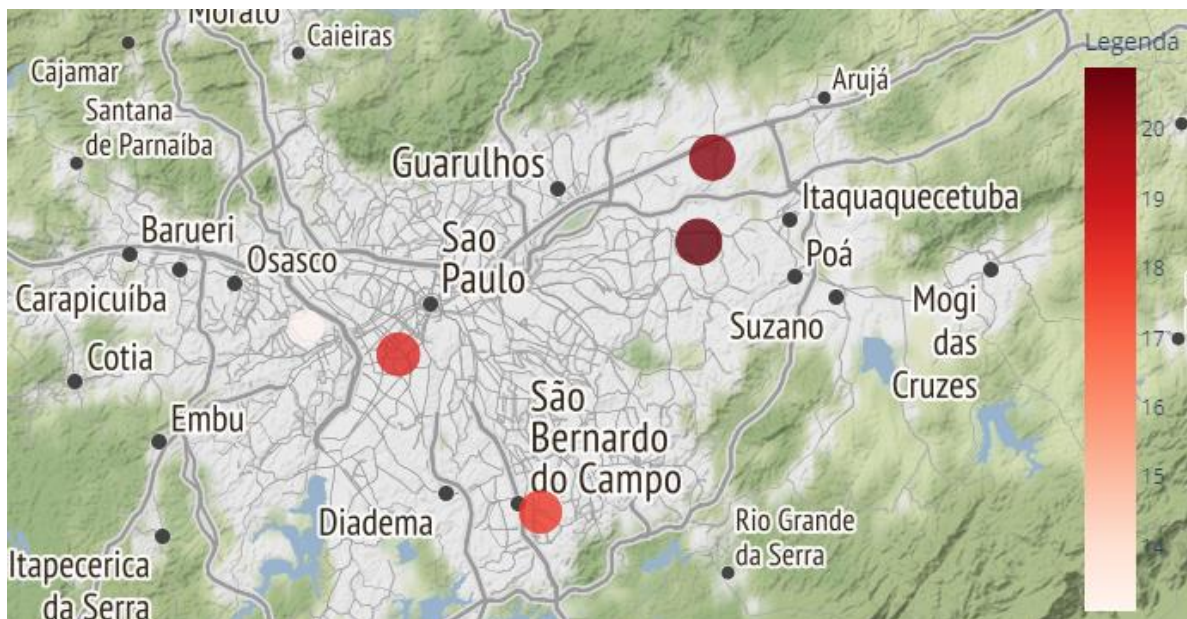


Figura 57 - Concentração média de $MP_{2.5}$, em $\mu g/m^3$, no período frio e seco e, 2016.

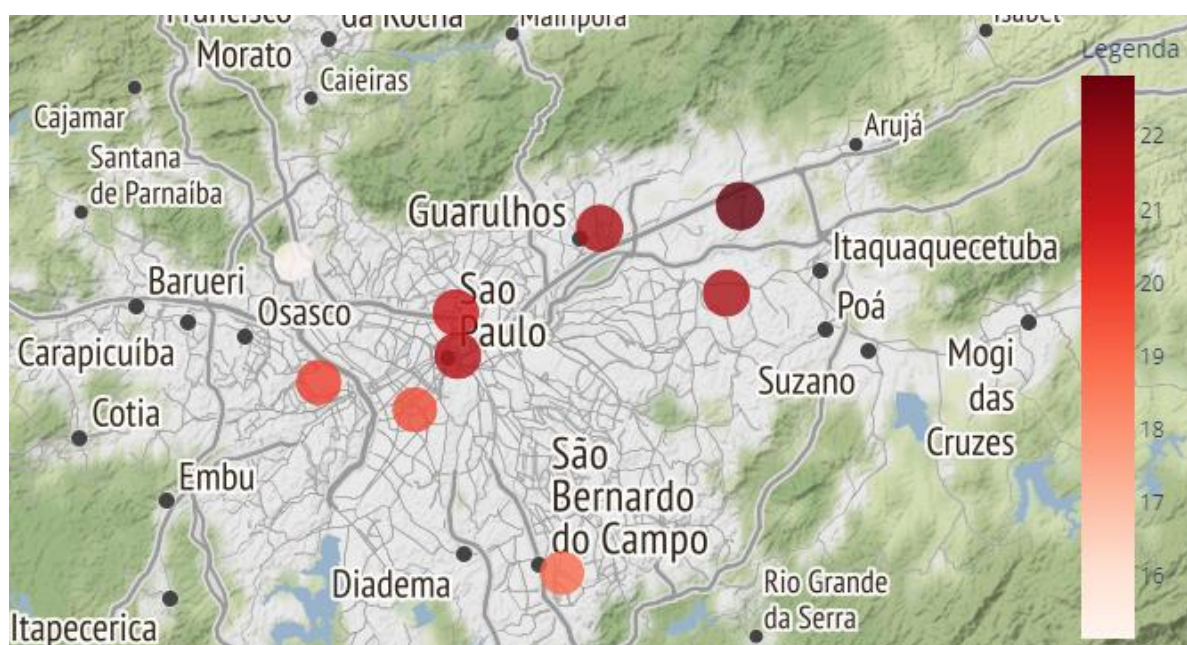


Figura 58 - Concentração média de $MP_{2.5}$, em $\mu g/m^3$, no período frio e seco e, 2017.

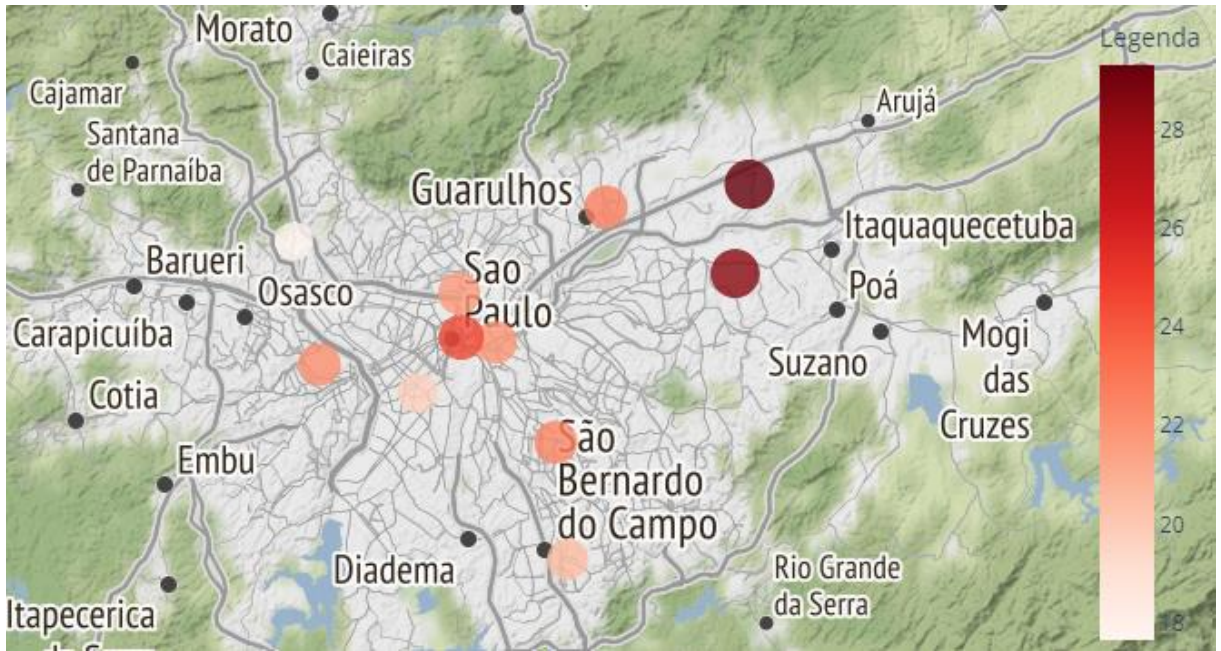


Figura 59 - Concentração média de MP_{2.5}, em µg/m³, no período frio e seco e, 2018.

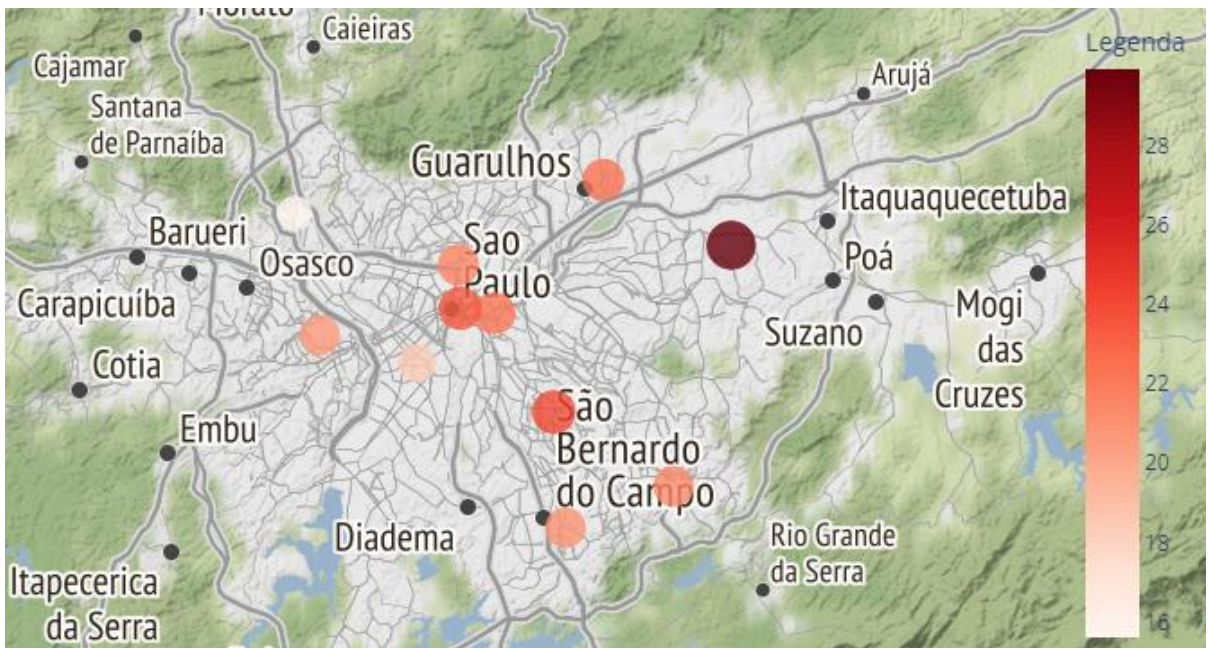


Figura 60 - Concentração média de MP_{2.5}, em µg/m³, no período frio e seco e, 2019.

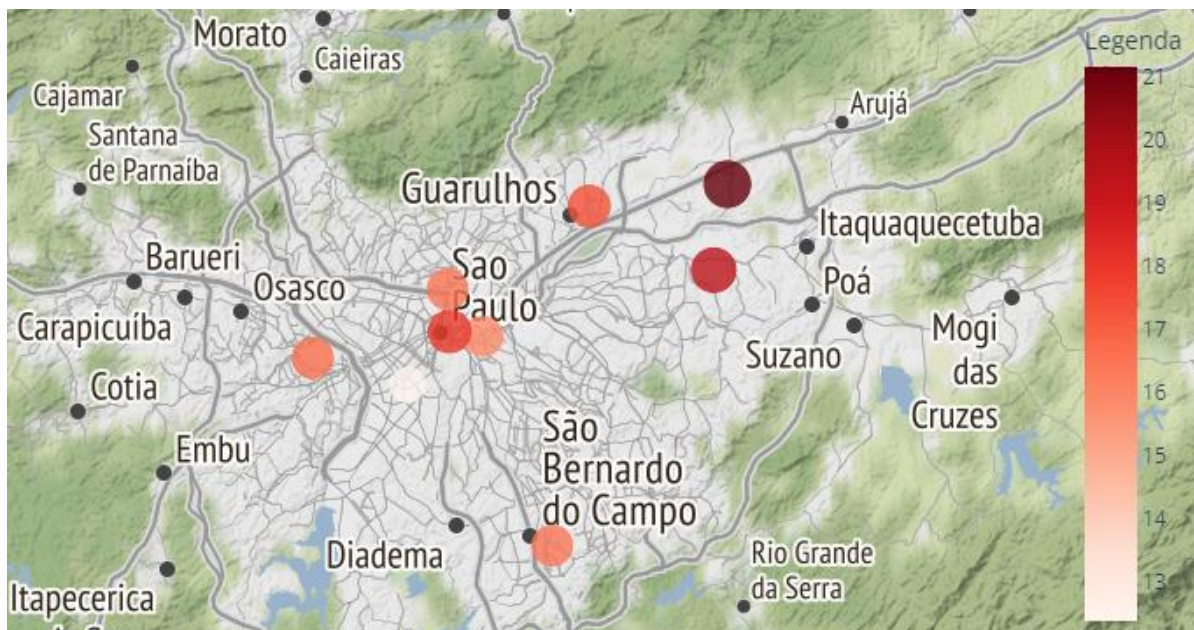


Figura 61 - Concentração média de MP_{2.5}, em µg/m³, no período frio e seco e, 2020.

Assim como para o poluente MP₁₀, a estação Guarulhos-Pimentas (bairro) apresenta as maiores médias para a maioria dos períodos analisados. Os períodos frios e secos novamente apresentam médias superiores aos períodos quentes e úmidos.

O ano de 2016 para todas as figuras apresenta poucos círculos, devido à quantidade significativa de dados inválidos para a maior parte das estações, que impossibilitou a sua utilização na análise.

A Tabela 10 apresenta os principais resultados das análises multivariadas para MP_{2.5}. A coluna MC mostra os pares de estações mais correlacionados na matriz de correlação, com coeficiente acima de 0,75. A porcentagem da coluna PCA indica a variância explicada pelo primeiro componente principal (PC1). Nesta coluna estão listadas também as estações com pesos no PC1 superiores a 0,35. A coluna AF mostra as estações com maiores *loadings* no primeiro fator, acima de 0,80 em número absoluto, além de mostrar também quantos fatores foram considerados para a análise. E, por fim, a coluna AC contém os pares de estações com maior grau de similaridade.

Tabela 10 - Resumo dos principais resultados das técnicas de análise estatística multivariada para o poluente MP_{2.5}.

Base de Dados	MC	PCA	AF	AC
2016	Sem correlação acima de 0,75.	54,8% Itaim Paulista S. Bernardo-	1 fator Itaim Paulista	Guarulhos-Pimentas - Itaim Paulista

		Centro Guarulhos- Pimentas Ibirapuera 57,3%		Ibirapuera - S. Bernardo- Centro
2016qu	Sem correlação acima de 0,75.	Cid. Universitária- USP-Ipen Ibirapuera Itaim Paulista S. Bernardo- Centro 57,5%	1 fator Sem loadings acima de 0,80.	Cid. Universitária- USP-Ipen - S. Bernardo- Centro
2016fs	Sem correlação acima de 0,75.	Itaim Paulista Guarulhos- Pimentas S. Bernardo- Centro Ibirapuera 60,8%	1 fator Itaim Paulista	Guarulhos- Pimentas - Itaim Paulista Ibirapuera - S. Bernardo- Centro
2017	Parque D. Pedro II - Santana	Santana Parque D. Pedro II Itaim Paulista Ibirapuera 61,0%	1 fator Ibirapuera Itaim Paulista Parque D. Pedro II Santana	Parque D. Pedro II - Santana Guarulhos- Pimentas - Itaim Paulista
2017qu	Sem correlação acima de 0,75.	Parque D. Pedro II Cid. Universitária- USP-Ipen Ibirapuera Pico do Jaraguá S. Bernardo- Centro 60,6%	1 fator Cid. Universitária- USP-Ipen Parque D. Pedro II	Cid. Universitária- USP-Ipen - Parque D. Pedro II
2017fs	Parque D. Pedro II - Santana	Parque D. Pedro II Santana Ibirapuera Itaim Paulista 66,4%	1 fator Ibirapuera Itaim Paulista Parque D. Pedro II Santana	Guarulhos- Pimentas - Itaim Paulista Parque D. Pedro II - Santana
2018	Mooca - Parque D. Pedro II	Sem pesos	1 fator Guarulhos-Paço	Mooca - Parque D. Pedro II

S. Bernardo- Centro - S Caetano	acima de 0,35.	Municipal Ibirapuera Mooca Parque D. Pedro II	S. Bernardo- Centro - S Caetano
Ibirapuera - Parque D. Pedro II		S. Bernardo- Centro S Caetano Santana	
Mooca - S Caetano			
Ibirapuera - Mooca			
Parque D. Pedro II - Santana			
Mooca - Santana			
Parque D. Pedro II - S. Caetano			
Mooca - S. Bernardo-Centro			
Parque D. Pedro II - S. Bernardo- Centro			

2018qu		47,6%		
	Sem correlação acima de 0,75.	S Caetano Santana Ibirapuera Parque D. Pedro II	2 fatores Sem loadings acima de 0,80.	Ibirapuera - S Caetano Parque D. Pedro II - Santana
2018fs	Mooca - Parque D. Pedro II		1 fator	Mooca - Parque D. Pedro II
	S. Bernardo- Centro - S Caetano	68,1%	Guarulhos-Paço Municipal Ibirapuera Mooca Parque D. Pedro II	S. Bernardo- Centro - S Caetano
	Mooca - S Caetano	Sem pesos acima de 0,35.	S. Bernardo- Centro S Caetano Santana	Cid. Universitária- USP-Ipen - Ibirapuera
	Parque D. Pedro II - Santana			
	Ibirapuera - Parque D. Pedro II			

Parque D. Pedro II
- S Caetano

Mooca - S.
Bernardo-Centro

Mooca - Santana

Ibirapuera -
Mooca

Parque D. Pedro II
- S. Bernardo-
Centro

2019	Mooca - Parque D. Pedro II		1 fator	
	Mooca - S Caetano	65,6%	Guarulhos-Paço Municipal	Mooca - Parque D. Pedro II
	Ibirapuera - Mooca	Mooca	Ibirapuera	S. Bernardo-Centro – S. Caetano
	Mooca - Santana	Parque D. Pedro II	Parque D. Pedro II	
	Ibirapuera - Parque D. Pedro II	S Caetano	S Caetano	
2019qu		50,3%	2 fatores	Ibirapuera - Mooca
	Ibirapuera - Mooca	Mooca	Guarulhos-Paço Municipal	Guarulhos-Paço Municipal - Itaim Paulista
2019fs	Mooca - Parque D. Pedro II		1 fator	
	Mooca - S Caetano		Guarulhos-Paço Municipal	Mooca - Parque D. Pedro II
	S. Bernardo-Centro - S Caetano	67,4%	Ibirapuera	
	Parque D. Pedro II - S Caetano	Sem pesos acima de 0,35.	Itaim Paulista	S. Bernardo-Centro-S Caetano
	Ibirapuera - Mooca		Mooca	

	Itaim Paulista - S Caetano			
	Mooca - S. Bernardo-Centro			
	Ibirapuera - Parque D. Pedro II			
	Itaim Paulista - Guarulhos-Paço Municipal			
	Guarulhos-Paço Municipal - S. Caetano			
	Cid. Universitária- USP-Ipen - Mooca			
	Mooca - Santana			
	Guarulhos- Pimentas - Itaim Paulista		1 fator	Guarulhos- Pimentas-Itaim Paulista
2020	Mooca - Parque D. Pedro II	68,2%	Cid. Universitária- USP-Ipen	Cid. Universitária- USP-Ipen - Mooca
	Cid. Universitária- USP-Ipen - Ibirapuera	Sem pesos acima de 0,35.	Itaim Paulista Mooca Parque D. Pedro II	Parque D. Pedro II - Santana
	Mooca - S Caetano		S Caetano Santana	
	Parque D. Pedro II - Santana			
2020qu		57,3%	1 fator	Mooca - Santana
	Sem correlação acima de 0,75.	Mooca Santana S Caetano Guarulhos-Paço Municipal	Mooca Santana	Guarulhos- Pimentas - Itaim Paulista
2020fs	Mooca - Parque D. Pedro II	72,6%	1 fator	Cid. Universitária- USP-Ipen - Ibirapuera
	Cid. Universitária-	Mooca Parque D. Pedro	Cid. Universitária-	

	USP-Ipen - Ibirapuera	II	USP-Ipen Guarulhos-Paço Municipal	Mooca - Parque D. Pedro II
	Cid. Universitária- USP-Ipen - Mooca		Ibirapuera Mooca Parque D. Pedro II	
	Ibirapuera - Mooca		Santana	
	Mooca - Santana			
	Cid. Universitária- USP-Ipen - Parque D. Pedro II			
	Parque D. Pedro II - Santana			
	Ibirapuera - Parque D. Pedro II			
		60,0%		
2016-2020	Parque D. Pedro II - Santana	Parque D. Pedro II	1 fator Ibirapuera	Parque D. Pedro II - Santana
	Ibirapuera – Parque D. Pedro II	Santana Ibirapuera Guarulhos-Paço Municipal	Parque D. Pedro II Santana	Guarulhos-Paço Municipal - Itaim Paulista

Para a maioria dos períodos considerados, a variância explicada pelo primeiro componente principal é inferior a 65%, principalmente nos anos de 2016 e 2017, que apresentaram uma quantidade significativa de dados inválidos, diminuindo as bases de dados analisadas. Nota-se que os períodos com maior precipitação (quente e úmido) apresentam baixas correlações, devido às condições meteorológicas, que contribuem para menores níveis desse poluente.

As estações com maiores pesos no primeiro componente principal na PCA, na maior parte dos períodos, são Ibirapuera (V4) e Parque D. Pedro II (V8), sendo estas as estações que mais contribuem para a variância do sistema no caso do MP_{2,5}.

No caso da análise de fatores, ao analisar as estações de monitoramento com maior carga para o primeiro fator, pode-se destacar para a maior parte dos períodos as estações Parque D. Pedro II (V8) e Santana (V12), que possuem escala de representatividade bairro e média, respectivamente. Já as estações com menores cargas no primeiro fator para a maioria

dos períodos são Pico do Jaraguá (V9), Cid. Universitária-USP-Ipen (V1) e Guarulhos-Pimentas (V3). Dessa forma, assim como ocorreu com os demais poluentes, uma possível interpretação do primeiro fator é que ele está associado à proximidade de vias de tráfego movimentadas.

Novamente, assim como ocorreu na análise de fatores do MP₁₀, há análises de períodos quentes e úmidos que apresentaram mais de um fator com autovalor superior a 1. No caso dos períodos “2018qu” e “2019qu”, a análise resultou em 2 fatores.

Assim como para o MP₁₀ e conforme mostrado nas tabelas do APÊNDICE G, não foi possível observar um padrão claro entre as cargas fatoriais do segundo fator. Logo, não foi possível chegar a conclusões satisfatórias a respeito da interpretação deste fator. De qualquer forma, acredita-se que esteja relacionado à presença de chuvas, assim como nas análises de MP₁₀.

As análises das matrizes de correlação e as análises de *clusters* para todos os períodos mostram que não há pares de estações que se repetem consistentemente nas análises como as mais correlacionadas. As que se repetem com maior frequência são: Parque D. Pedro II (V8) e Santana (V12); Guarulhos-Pimentas (V3) e Itaim Paulista (V5); e Mooca (V7) e Parque D. Pedro II (V8). Mas estes pares não se repetem o suficiente para a maioria dos períodos, e não é possível concluir que elas poderiam ter medições redundantes, conforme mostra a Figura 62.

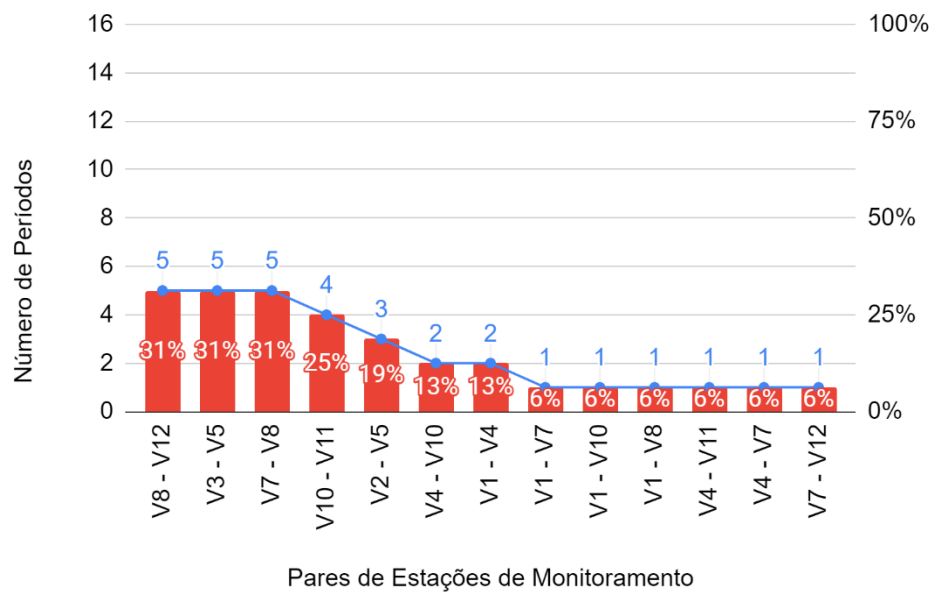


Figura 62 - Número de períodos para cada par de estações de monitoramento de MP_{2.5} com alto grau de similaridade na Análise de Clusters

Conclui-se que, no caso do poluente MP_{2.5}, não há evidências, a partir das análises multivariadas das bases de dados disponíveis, de que existem estações altamente correlacionadas, a ponto de possibilitar uma diminuição no número dos analisadores.

5. CONCLUSÕES

Os poluentes avaliados apresentam diferentes resultados devido às suas diferentes naturezas físicas e mecanismos de formação. Para o ozônio (O_3), poluente secundário formado na atmosfera, as correlações entre as estações medidoras e os níveis de similaridade dos agrupamentos foram superiores às dos materiais particulados (MP_{10}), e materiais particulados finos ($MP_{2.5}$), poluentes primários emitidos diretamente por fontes como veículos automotores, responsáveis pela maior parte das emissões da RMSP. Além da diferença entre a formação dos poluentes, há variações entre os períodos do ano para os quais foram feitas as análises. No caso de MP_{10} e $MP_{2.5}$, as correlações entre as estações nos períodos quentes e úmidos são inferiores às dos períodos frios e secos, o que revela a influência de condições meteorológicas nas concentrações desses poluentes na atmosfera.

No estudo, a análise de clusters se mostrou mais potente que a matriz de correlação, pois ela oferece uma hierarquia de similaridade entre as estações de monitoramento de concentrações de poluentes. As análises de componentes principais e de fatores também se mostraram coerentes, sendo as estações com os maiores pesos no primeiro componente principal as mesmas com as maiores cargas no primeiro fator.

A aplicação dos diferentes métodos de análise multivariada apresentou resultados para a identificação de semelhanças entre estações de monitoramento da CETESB para os poluentes atmosféricos considerados neste estudo. No caso do poluente O_3 , devido às altas correlações entre as estações, é difícil concluir se existem medidores de fato redundantes. Esse resultado é coerente com o fenômeno físico-químico de formação deste poluente, uma vez que ela depende da luz, que geralmente é distribuída de forma uniforme na RMSP. O par de estações que se destacou pelas altas correlações de suas medições é Mooca e Parque D. Pedro II, que possuem escala de representatividade bairro e estão distantes em aproximadamente 2,8 km uma da outra.

Para o MP_{10} , as estações Capão Redondo e Santo Amaro apresentaram maior potencial de terem um de seus analisadores removidos, sendo que Santo Amaro representa a maior parte da variância do sistema. Quanto ao $MP_{2.5}$, devido ao menor número de estações monitoras na maior parte das análises, não é possível observar tendências de correlação entre as estações ao longo dos anos analisados, o que dificulta a indicação de estações redundantes.

Com este estudo preliminar, é possível que a CETESB investigue de uma maneira mais direcionada as possibilidades de realocação dos analisadores mencionados. Além disso, sugere-se o seguinte para estudos de continuação:

- Antes de aplicar as técnicas multivariadas, ainda no tratamento de dados, em vez de eliminar os dados inválidos, testar outras técnicas como imputação de dados (*data imputation*) para verificar se são obtidos resultados mais satisfatórios, principalmente nos poluentes O₃ e MP_{2,5}.
- Realizar análises adicionais de fatores nos períodos quentes e úmidos nas bases de dados de material particulado, tanto MP₁₀ quanto MP_{2,5}, com o intuito de encontrar a interpretação de outros fatores não mensuráveis que possam influenciar as concentrações medidas pelas estações de monitoramento.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; DE MORAES GONÇALVES, J. L.; SPAROVEK, G. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711–728, 2013.
- ANDRADE, M. de F.; KUMAR, P.; DE FREITAS, E. D.; YNOUE, R. Y.; MARTINS, J.; MARTINS, L. D.; NOGUEIRA, T.; PEREZ-MARTINEZ, P.; DE MIRANDA, R. M.; ALBUQUERQUE, T.; GONÇALVES, F. L. T.; OYAMA, B.; ZHANG, Y. Air quality in the megacity of São Paulo: Evolution over the last 30 years and future perspectives. **Atmospheric Environment**, v. 159, p. 66–82, 2017. Disponível em: <<http://dx.doi.org/10.1016/j.atmosenv.2017.03.051>>.
- CETESB. Relatorio-de-Qualidade-do-Ar-no-Estado-de-Sao-Paulo-2021. 2021.
- CETESB - COMPANHIA AMBIENTAL DO ESTADO DE SÃO PAULO. Classificação Exedita Da Representatividade Espacial Das Estações De Monitoramento Da Qualidade Do Ar Da Cetesb No Estado De São Paulo. p. 29, 2016.
- CONAMA. RESOLUÇÃO CONAMA nº 303, de 20 de março de 2002. **DOU no. 224**, p. 90–92, 2002.
- CONAMA. RESOLUÇÃO CONAMA Nº 491 DE 19/11/2018. 2018.
- CORTEZ, L. A. B. **Proálcool 40 anos**. [s.l: s.n.]v. 161–153 p.
- DE FATIMA ANDRADE, M.; DE MIRANDA, R. M.; FORNARO, A.; KERR, A.; OYAMA, B.; DE ANDRE, P. A.; SALDIVA, P. Vehicle emissions and PM 2.5 mass concentrations in six Brazilian cities. **Air Quality, Atmosphere and Health**, v. 5, n. 1, p. 79–88, 2012.
- DE PAULA, R. R. R. Análise comparativa de modelos de estatística multivariada aplicados à previsão de níveis de poluentes atmosféricos. **Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia Química**, p. 97, 2017.
- FRANÇA, A. L. F. **GUIA TÉCNICO PARA O MONITORAMENTO E AVALIAÇÃO DA QUALIDADE DO AR**. Ministério do Meio Ambiente Secretaria de Qualidade Ambiental, 2019.
- FUJIKOSHI, Y. TESTS FOR REDUNDANCY OF SOME VARIABLES IN MULTIVARIATE ANALYSIS. **Statistical Data Analysis and Inference**, p. 141–163, 1 jan. 1989. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B9780444880291500186>>. Acesso em: 15 fev. 2020.
- FUJIKOSHI, Y.; KHATRI, C. G. A study of redundancy of some variables in covariate discriminant analysis. **Annals of the Institute of Statistical Mathematics**, v. 42, n. 4, p. 769–782, 1990.

GLADWELL, M. Fora de Série - Outliers. **GMT Editores Ltda**, v. 13, n. 1, p. 104–116, 2008.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J.. The elements of statistical learning. **Springer Texts**, p. 764, 2008.

PIMENTEL, A.; FILHO, V. Otimização das Estações de Monitoramento da CETESB através de Técnicas de Análise Multivariada. **TCC**, n. April, p. 98, 2019.

RUDKE, A. P.; MARTINS, J. A.; DE ALMEIDA, D. S.; MARTINS, L. D.; BEAL, A.; HALLAK, R.; FREITAS, E. D.; ANDRADE, M. F.; FOROUTAN, H.; BAEK, B. H.; DE, T. T. How mobility restrictions policy and atmospheric conditions impacted air quality in the State of São Paulo during the COVID-19 outbreak. **Environmental Research**, v. 198, n. April, p. 111255, 2021. Disponível em: <<https://doi.org/10.1016/j.envres.2021.111255>>.

SILVA, M. A. Da; SANTOS, J.; COSTA, M. C.; RIBEIRO, L.; PROCESSOS, M. D. E.; EM, D. E. P. Análise estatística multivariada aplicada a dados hidroquímicos. **6º Congresso da Água**, p. 2, 2002. Disponível em: <<http://hdl.handle.net/10451/10457>>.

YUAN, G. L.; SUN, T. H.; HAN, P.; LI, J. Environmental geochemical mapping and multivariate geostatistical analysis of heavy metals in topsoils of a closed steel smelter: Capital Iron & Steel Factory, Beijing, China. **Journal of Geochemical Exploration**, v. 130, p. 15–21, 2013. Disponível em: <<http://dx.doi.org/10.1016/j.gexplo.2013.02.010>>.

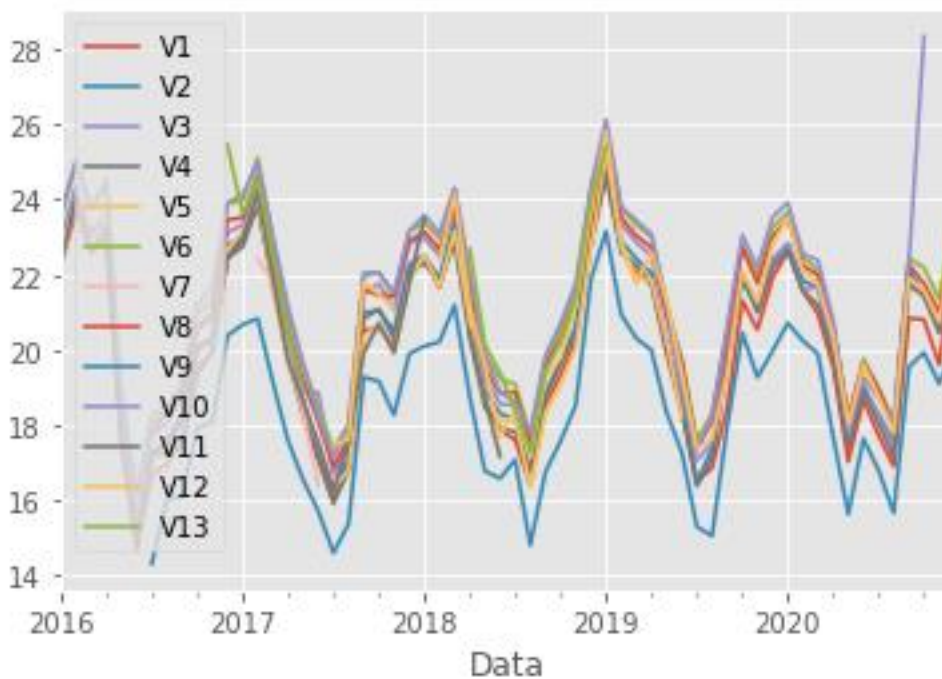
APÊNDICE A – Seleção de períodos pela temperatura e umidade relativa

Figura 63 - Dados de temperatura (°C) no período de 2016 a 2020.

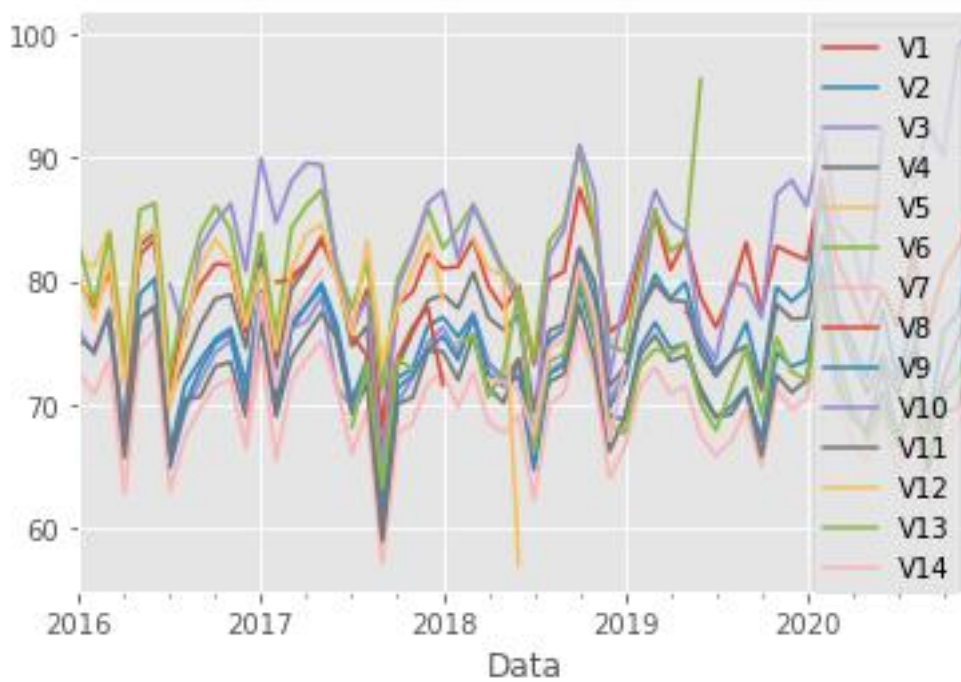


Figura 64 - Dados de umidade relativa do ar (%) no período de 2016 a 2020.

APÊNDICE B – Exemplos de matrizes de correlação

Tabela 11 – Matriz de correlação do poluente O₃ no período de 2016 a 2020.

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V13	V14	V15	V16	V17	V18	V19	V20	V21
V1	1.000	0.902	0.877	0.913	0.807	0.842	0.825	0.904	0.939	0.844	0.819	0.856	0.843	0.889	0.613	0.860	0.877	0.898	0.856	0.937
V2	0.902	1.000	0.918	0.865	0.811	0.874	0.849	0.917	0.879	0.849	0.764	0.893	0.869	0.920	0.667	0.813	0.818	0.877	0.884	0.909
V3	0.877	0.918	1.000	0.871	0.867	0.874	0.839	0.947	0.892	0.858	0.762	0.917	0.880	0.915	0.668	0.825	0.821	0.887	0.910	0.889
V4	0.913	0.865	0.871	1.000	0.833	0.857	0.840	0.915	0.951	0.856	0.860	0.880	0.856	0.908	0.633	0.914	0.932	0.946	0.876	0.932
V5	0.807	0.811	0.867	0.833	1.000	0.852	0.838	0.874	0.848	0.883	0.785	0.884	0.806	0.851	0.596	0.831	0.814	0.867	0.859	0.817
V6	0.842	0.874	0.874	0.857	0.852	1.000	0.930	0.901	0.861	0.920	0.794	0.885	0.906	0.908	0.677	0.856	0.824	0.884	0.935	0.845
V7	0.825	0.849	0.839	0.840	0.838	0.930	1.000	0.871	0.840	0.944	0.812	0.863	0.848	0.877	0.597	0.859	0.818	0.873	0.879	0.831
V8	0.904	0.917	0.947	0.915	0.874	0.901	0.871	1.000	0.928	0.889	0.814	0.939	0.897	0.954	0.665	0.882	0.877	0.939	0.929	0.922
V9	0.939	0.879	0.892	0.951	0.848	0.861	0.840	0.928	1.000	0.864	0.859	0.878	0.857	0.899	0.623	0.897	0.924	0.931	0.880	0.943
V10	0.844	0.849	0.858	0.856	0.883	0.920	0.944	0.889	0.864	1.000	0.834	0.886	0.836	0.891	0.588	0.874	0.843	0.895	0.880	0.853
V11	0.819	0.764	0.762	0.860	0.785	0.794	0.812	0.814	0.859	0.834	1.000	0.774	0.773	0.811	0.511	0.914	0.882	0.874	0.786	0.822
V13	0.856	0.893	0.917	0.880	0.884	0.885	0.863	0.939	0.878	0.886	0.774	1.000	0.874	0.953	0.643	0.841	0.842	0.916	0.921	0.886
V14	0.843	0.869	0.880	0.856	0.806	0.906	0.848	0.897	0.857	0.836	0.773	0.874	1.000	0.910	0.742	0.837	0.814	0.862	0.960	0.833
V15	0.889	0.920	0.915	0.908	0.851	0.908	0.877	0.954	0.899	0.891	0.811	0.953	0.910	1.000	0.672	0.877	0.862	0.935	0.942	0.913
V16	0.613	0.667	0.668	0.633	0.596	0.677	0.597	0.665	0.623	0.588	0.511	0.643	0.742	0.672	1.000	0.589	0.571	0.614	0.716	0.617
V17	0.860	0.813	0.825	0.914	0.831	0.856	0.859	0.882	0.897	0.874	0.914	0.841	0.837	0.877	0.589	1.000	0.913	0.936	0.857	0.862
V18	0.877	0.818	0.821	0.932	0.814	0.824	0.818	0.877	0.924	0.843	0.882	0.842	0.814	0.862	0.571	0.913	1.000	0.919	0.830	0.881
V19	0.898	0.877	0.887	0.946	0.867	0.884	0.873	0.939	0.931	0.895	0.874	0.916	0.862	0.935	0.614	0.936	0.919	1.000	0.896	0.923
V20	0.856	0.884	0.910	0.876	0.859	0.935	0.879	0.929	0.880	0.880	0.786	0.921	0.960	0.942	0.716	0.857	0.830	0.896	1.000	0.858
V21	0.937	0.909	0.889	0.932	0.817	0.845	0.831	0.922	0.943	0.853	0.822	0.886	0.833	0.913	0.617	0.862	0.881	0.923	0.858	1.000

Tabela 12 - Matriz de correlação do poluente MP₁₀ no período de 2016 a 2020.

	V1	V2	V3	V4	V5	V7	V8	V9	V12	V13	V14	V15	V16	V18
V1	1.000	0.782	0.653	0.665	0.519	0.745	0.717	0.586	0.697	0.757	0.582	0.693	0.744	0.850
V2	0.782	1.000	0.610	0.698	0.548	0.690	0.728	0.559	0.694	0.725	0.571	0.658	0.693	0.750
V3	0.653	0.610	1.000	0.607	0.498	0.622	0.632	0.563	0.569	0.652	0.545	0.642	0.683	0.670
V4	0.665	0.698	0.607	1.000	0.664	0.661	0.765	0.616	0.750	0.755	0.619	0.664	0.723	0.670
V5	0.519	0.548	0.498	0.664	1.000	0.510	0.649	0.503	0.533	0.610	0.515	0.550	0.589	0.532
V7	0.745	0.690	0.622	0.661	0.510	1.000	0.682	0.606	0.654	0.730	0.607	0.676	0.724	0.783
V8	0.717	0.728	0.632	0.765	0.649	0.682	1.000	0.636	0.688	0.752	0.665	0.683	0.749	0.703
V9	0.586	0.559	0.563	0.616	0.503	0.606	0.636	1.000	0.568	0.619	0.606	0.600	0.682	0.622
V12	0.697	0.694	0.569	0.750	0.533	0.654	0.688	0.568	1.000	0.749	0.564	0.651	0.701	0.683
V13	0.757	0.725	0.652	0.755	0.610	0.730	0.752	0.619	0.749	1.000	0.611	0.719	0.792	0.768
V14	0.582	0.571	0.545	0.619	0.515	0.607	0.665	0.606	0.564	0.611	1.000	0.599	0.661	0.582
V15	0.693	0.658	0.642	0.664	0.550	0.676	0.683	0.600	0.651	0.719	0.599	1.000	0.759	0.710
V16	0.744	0.693	0.683	0.723	0.589	0.724	0.749	0.682	0.701	0.792	0.661	0.759	1.000	0.768
V18	0.850	0.750	0.670	0.670	0.532	0.783	0.703	0.622	0.683	0.768	0.582	0.710	0.768	1.000

Tabela 13 - Matriz de correlação do poluente MP_{2,5} no período de 2016 a 2020.

	V1	V2	V3	V4	V5	V8	V9	V10	V12
V1	1.000	0.522	0.364	0.668	0.553	0.659	0.343	0.571	0.597
V2	0.522	1.000	0.546	0.602	0.667	0.672	0.350	0.552	0.693
V3	0.364	0.546	1.000	0.437	0.559	0.496	0.238	0.434	0.497
V4	0.668	0.602	0.437	1.000	0.592	0.744	0.376	0.644	0.701
V5	0.553	0.667	0.559	0.592	1.000	0.670	0.310	0.595	0.614
V8	0.659	0.672	0.496	0.744	0.670	1.000	0.325	0.681	0.771
V9	0.343	0.350	0.238	0.376	0.310	0.325	1.000	0.302	0.385
V10	0.571	0.552	0.434	0.644	0.595	0.681	0.302	1.000	0.620
V12	0.597	0.693	0.497	0.701	0.614	0.771	0.385	0.620	1.000

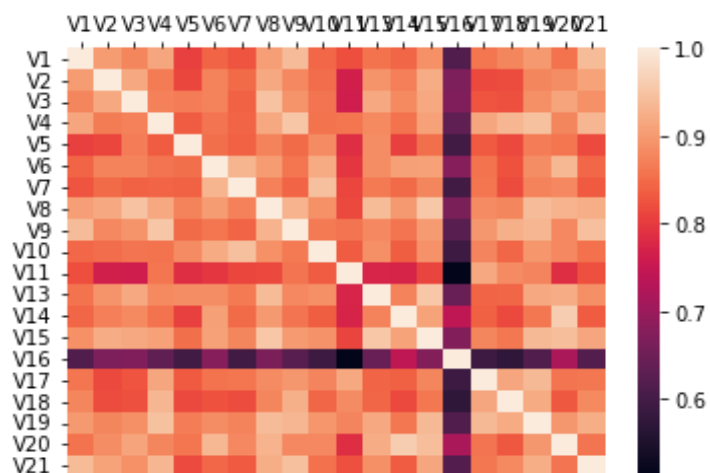


Figura 65 - Representação gráfica da matriz de correlação do poluente O₃ no período de 2016 a 2020.

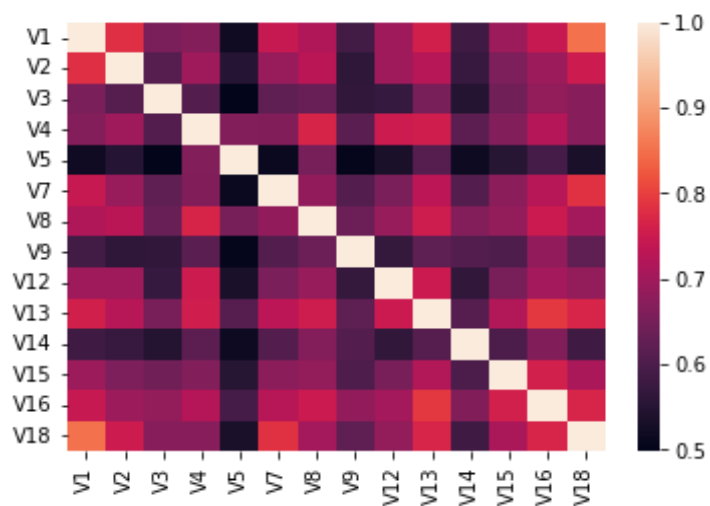


Figura 66 - Representação gráfica da matriz de correlação do poluente MP₁₀ no período de 2016 a 2020.

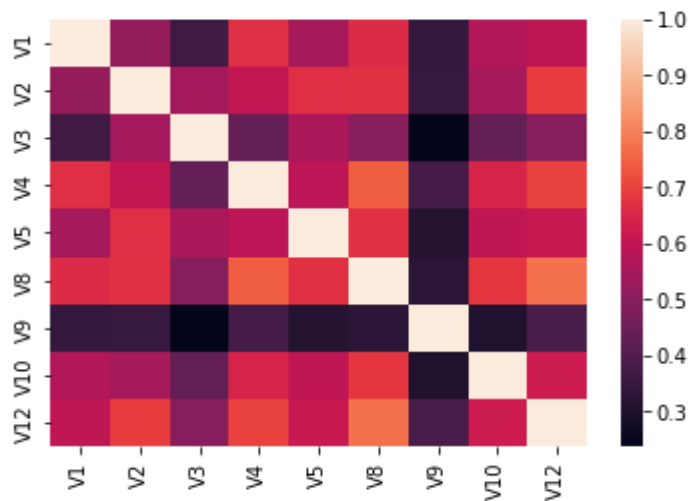


Figura 67 - Representação gráfica da matriz de correlação do poluente MP_{2.5} no período de 2016 a 2020.

APÊNDICE C – Exemplos de análises de componentes principais (PCA)

a) Ozônio (O₃)

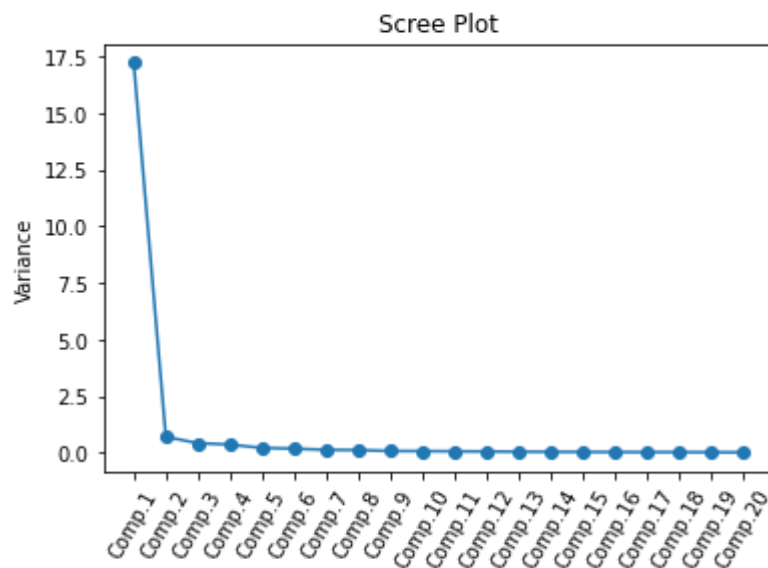


Figura 68 - Scree plot da PCA do poluente O₃ no período de 2016 a 2020.

Tabela 14 - Tabela resumo da PCA do poluente O₃ no período de 2016 a 2020.

	Desvio Padrão	Proporção da Variância	Proporção Cumulativa
PC1	4.148492	0.860499	0.860499
PC2	0.843368	0.035564	0.896063
PC3	0.655091	0.021457	0.91752
PC4	0.608956	0.018541	0.936061
PC5	0.468	0.010951	0.947012
PC6	0.437396	0.009566	0.956578
PC7	0.366627	0.006721	0.963299
PC8	0.356452	0.006353	0.969652
PC9	0.291501	0.004249	0.9739
PC10	0.285487	0.004075	0.977976
PC11	0.266512	0.003551	0.981527
PC12	0.25189	0.003172	0.984699
PC13	0.231921	0.002689	0.987389
PC14	0.216657	0.002347	0.989736
PC15	0.207211	0.002147	0.991883
PC16	0.194991	0.001901	0.993784
PC17	0.190562	0.001816	0.995599
PC18	0.185939	0.001729	0.997328
PC19	0.16994	0.001444	0.998772
PC20	0.156714	0.001228	1

Tabela 15 – Pesos das variáveis sobre o componente 1, em ordem decrescente, do poluente O₃ no período de 2016 a 2020.

Variáveis	PC1
V8	0.233957
V15	0.232926
V19	0.232846
V9	0.230463
V4	0.229731
V20	0.229701
V13	0.228331
V21	0.227659
V3	0.22678
V6	0.226724
V1	0.225382
V10	0.22513
V2	0.224927
V17	0.224476
V14	0.22363
V18	0.222264
V7	0.221812
V5	0.217702
V11	0.211632
V16	0.167591

b) Partículas inaláveis e fumaça (MP₁₀)

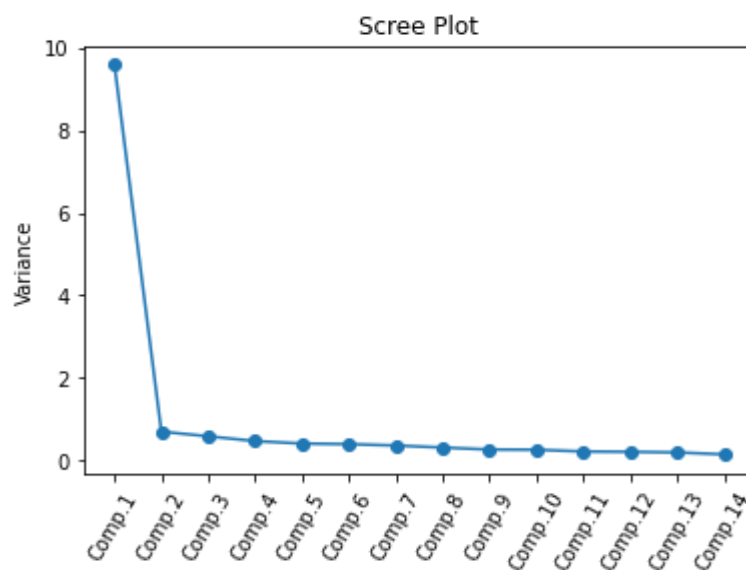


Figura 69 - Scree plot do PCA do poluente MP₁₀ no período de 2016 a 2020.

Tabela 16 - Tabela resumo da PCA do poluente MP₁₀ no período de 2016 a 2020.

	Desvio Padrão	Proporção da Variância	Proporção Cumulativa
PC1	3.09643	0.684848	0.684848
PC2	0.830912	0.049315	0.734164
PC3	0.759611	0.041215	0.775379
PC4	0.678709	0.032903	0.808282
PC5	0.633381	0.028655	0.836937
PC6	0.622003	0.027635	0.864572
PC7	0.594535	0.025248	0.88982
PC8	0.551443	0.021721	0.91154
PC9	0.505588	0.018259	0.929799
PC10	0.500898	0.017921	0.94772
PC11	0.456107	0.01486	0.96258
PC12	0.4472	0.014285	0.976865
PC13	0.430836	0.013259	0.990123
PC14	0.371852	0.009877	1

Tabela 17 - Pesos das variáveis sobre o componente 1, em ordem decrescente, do poluente MP₁₀ no período de 2016 a 2020.

Variáveis	PC1
V16	0.287274
V13	0.286824
V18	0.282875
V8	0.280751
V1	0.280173
V4	0.275254
V2	0.271759
V7	0.271285
V15	0.268433
V12	0.265881
V3	0.249271
V9	0.243571
V14	0.242346
V5	0.227403

c) Partículas inaláveis finas (MP_{2,5})

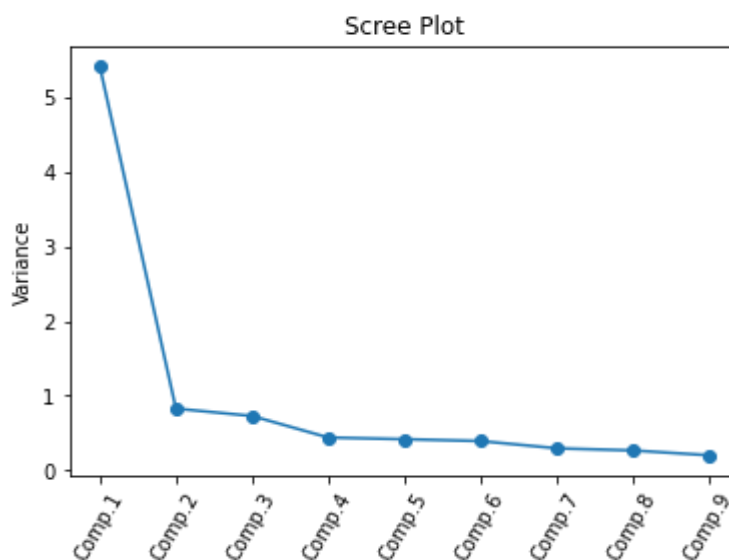


Figura 70 - Scree plot do PCA do poluente MP_{2,5} no período de 2016 a 2020.

Tabela 18 - Tabela resumo da PCA do poluente MP_{2,5} no período de 2016 a 2020.

	Desvio Padrão	Proporção da Variância	Proporção Cumulativa
PC1	2.326105	0.601196	0.601196
PC2	0.91055	0.092122	0.693318
PC3	0.854813	0.081189	0.774508
PC4	0.664975	0.049132	0.82364
PC5	0.647964	0.046651	0.870291
PC6	0.628573	0.0439	0.914192
PC7	0.545975	0.033121	0.947313
PC8	0.51894	0.029922	0.977235
PC9	0.452646	0.022765	1

Tabela 19 - Pesos das variáveis sobre o componente 1, em ordem decrescente, do poluente MP_{2,5} no período de 2016 a 2020.

Variáveis	PC1
V8	0.380131
V12	0.369243
V4	0.362013
V2	0.350366
V5	0.347735
V10	0.33865
V1	0.329769
V3	0.279404
V9	0.207202

APÊNDICE D – Exemplos de análises de fatores

a) Ozônio (O₃)

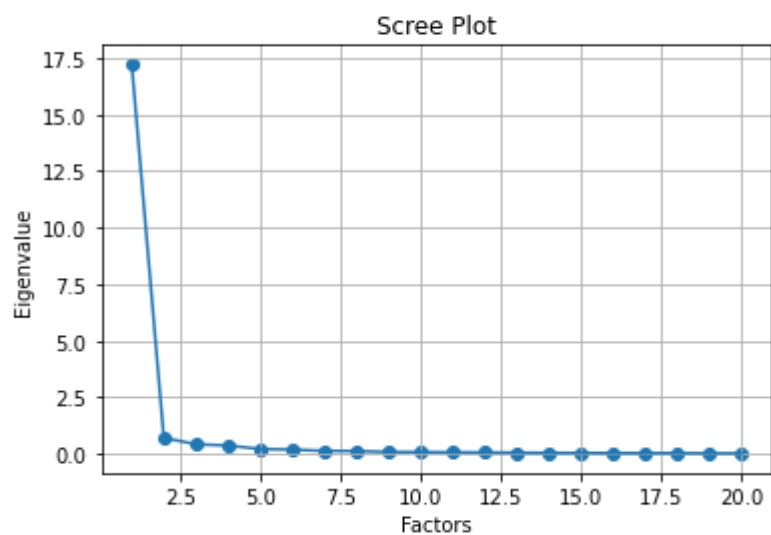


Figura 71 - Scree plot da análise de fatores do poluente O₃ no período de 2016 a 2020.

Tabela 20 - Cargas fatoriais do poluente O₃ no período de 2016 a 2020 em ordem decrescente.

	Fator 1
V8	0.971197
V15	0.966381
V19	0.966036
V9	0.954949
V4	0.951548
V20	0.951329
V13	0.945084
V21	0.941999
V3	0.937896
V6	0.937581
V1	0.931493
V10	0.93031
V2	0.929363
V17	0.927292
V14	0.923311
V18	0.917189
V7	0.915083
V5	0.896351
V11	0.868929
V16	0.676001

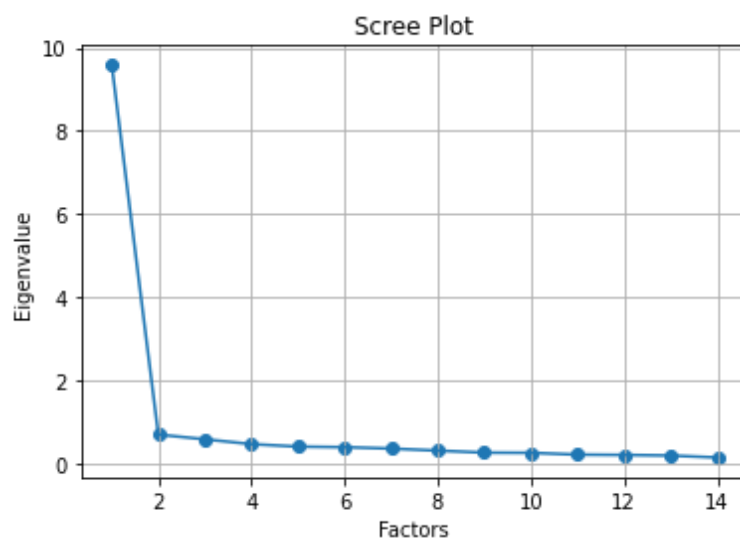
b) Partículas inaláveis e fumaça (MP₁₀)

Figura 72 - Scree plot da análise de fatores do poluente MP₁₀ no período de 2016 a 2020.

Tabela 21 - Cargas fatoriais do poluente MP₁₀ no período de 2016 a 2020 em ordem decrescente.

	Fator 1
V16	0.884343
V13	0.882861
V18	0.868504
V8	0.860247
V1	0.858713
V4	0.840316
V2	0.828245
V7	0.826502
V15	0.81618
V12	0.807331
V3	0.749394
V9	0.729768
V14	0.725571
V5	0.675741

c) Partículas inaláveis finas (MP_{2,5})

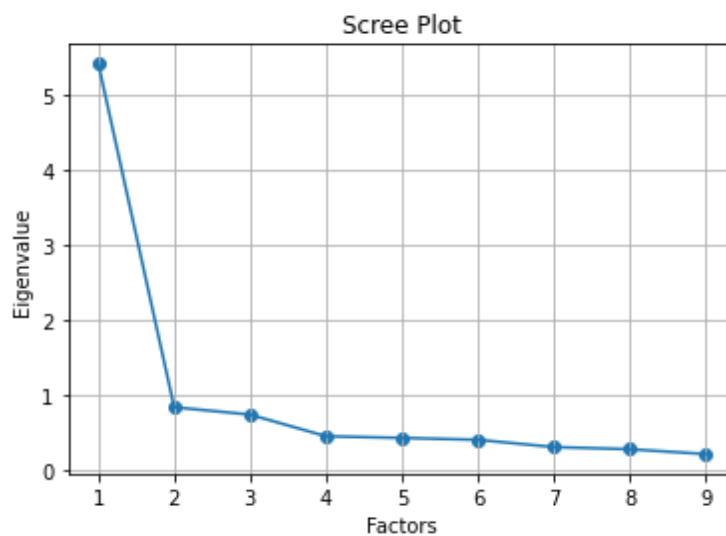


Figura 73 - Scree plot da análise de fatores do poluente MP_{2,5} no período de 2016 a 2020.

Tabela 22 - Cargas fatoriais do poluente MP_{2,5} no período de 2016 a 2020 em ordem decrescente.

	Fator 1
V8	0.881019
V12	0.844989
V4	0.822337
V2	0.786097
V5	0.778068
V10	0.752932
V1	0.727776
V3	0.594315
V9	0.424668

APÊNDICE E – Exemplos de análises de agrupamentos (*clusters*)

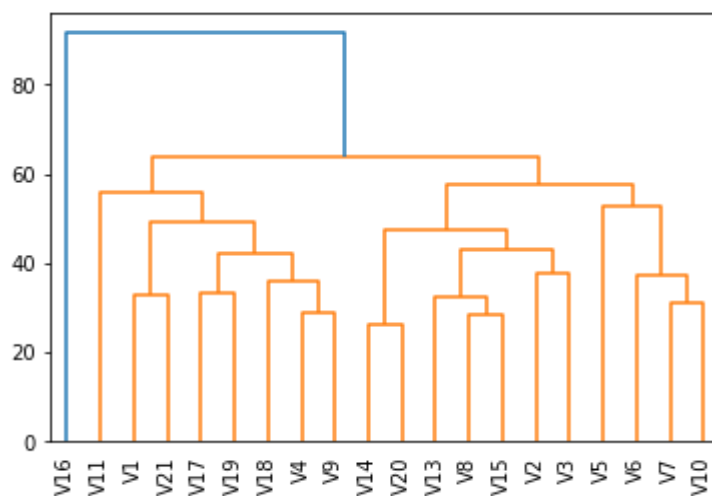


Figura 74 - Dendrograma da análise de clusters do poluente O₃ no período de 2016 a 2020.

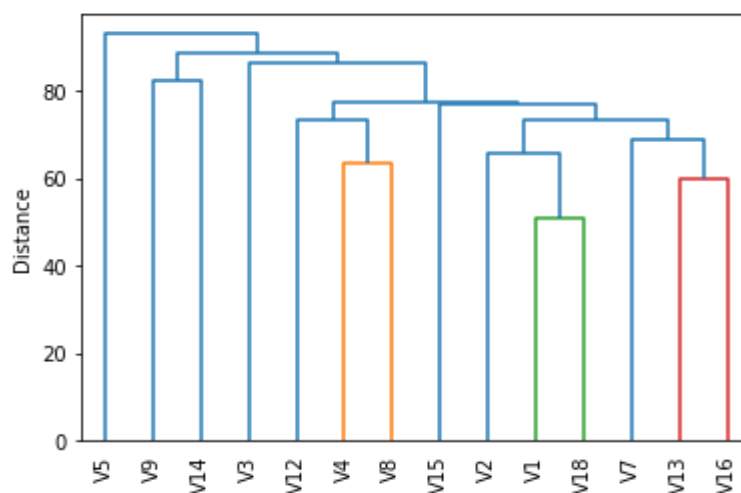


Figura 75 - Dendrograma da análise de clusters do poluente MP₁₀ no período de 2016 a 2020.

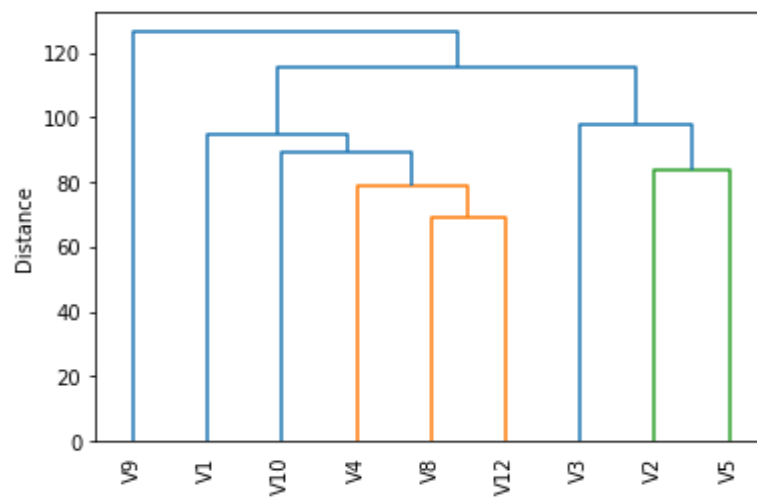


Figura 76 - Dendrograma da análise de clusters do poluente MP_{2,5} no período de 2016 a 2020.

APÊNDICE F – Cargas fatoriais das análises de fatores do poluente MP₁₀, períodos quentes e úmidos com mais de 1 fator

Tabela 23 - Cargas fatoriais do poluente MP₁₀, período quente e úmido do ano de 2016.

	Fator 1	Fator 2	Fator 3
V1	0.51515	-0.06461	0.126874
V2	0.380148	-0.24764	0.264148
V3	0.665796	0.01397	-0.15839
V4	0.623333	0.201574	-0.02569
V7	0.620697	0.064127	0.008792
V8	0.533289	0.624033	-0.02099
V9	0.664395	-0.2084	-0.17173
V12	0.38301	-0.20755	0.384642
V13	0.597716	0.582196	0.268918
V14	0.661758	0.243382	-0.3333
V15	0.716405	-0.35755	-0.02666
V16	0.724074	-0.27714	-0.20538
V17	0.671965	-0.01773	0.23102
V18	0.758278	-0.26702	0.000811

Tabela 24 - Cargas fatoriais do poluente MP₁₀, período quente e úmido do ano de 2017.

	Fator 1	Fator 2
V1	0.706517	0.254929
V2	0.373188	0.097164
V3	0.589033	0.170456
V4	0.691492	-0.20337
V5	0.71963	-0.37637
V7	0.696517	0.096936
V8	0.501923	-0.21274
V9	0.354261	0.022254
V12	0.74994	-0.34614
V13	0.852544	-0.11151
V14	0.64893	-0.00618
V15	0.757138	0.119389
V16	0.77585	0.155409
V18	0.800949	0.341807

Tabela 25 - Cargas fatoriais do poluente MP₁₀, período quente e úmido do ano de 2018.

	Fator 1	Fator 2
V1	0.804799	0.005482
V2	0.539216	-0.15456
V4	0.6522	0.040491
V7	0.573066	-0.07665
V8	0.801541	0.500049

V9	0.662175	0.144187
V12	0.703147	0.056648
V13	0.792729	-0.24819
V14	0.505052	0.397096
V15	0.75028	-0.13288
V16	0.793621	-0.25933
V18	0.791713	-0.17427

Tabela 26 - Cargas fatoriais do poluente MP₁₀, período quente e úmido do ano de 2019.

	Fator 1	Fator 2	Fator 3
V1	0.800239	0.352686	0.197913
V2	0.42764	0.132779	0.288888
V3	0.54123	0.09181	-0.12707
V4	0.754215	-0.39258	0.251945
V5	0.685988	-0.40697	0.16408
V7	0.725644	0.158397	0.084514
V8	0.798676	-0.16201	-0.00023
V9	0.613798	-0.16768	-0.36977
V12	0.76572	-0.30463	0.030396
V13	0.801504	-0.02182	0.114602
V14	0.7467	0.000124	-0.40019
V15	0.72479	0.071643	-0.16028
V16	0.805952	0.19595	-0.12601
V18	0.652219	0.528574	0.078986

Tabela 27 - Cargas fatoriais do poluente MP₁₀, período quente e úmido do ano de 2020.

	Fator 1	Fator 2
V1	0.771675	0.105592
V2	0.53727	-0.02074
V3	0.527913	0.031733
V4	0.794776	-0.17002
V5	0.584994	-0.10654
V7	0.532787	0.293256
V8	0.780998	-0.30053
V9	0.827105	-0.10751
V12	0.65102	0.03395
V13	0.567113	0.267198
V14	0.622095	-0.40435
V15	0.598316	-0.08219
V16	0.811287	0.022324
V18	0.856888	0.451309

APÊNDICE G – Cargas fatoriais das análises de fatores do poluente MP_{2,5}, períodos quentes e úmidos com mais de 1 fator

Tabela 28 - Cargas fatoriais do poluente MP_{2,5}, período quente e úmido do ano de 2018.

	Fator 1	Fator 2
V1	0.600712	-0.03948
V2	0.559403	0.343246
V3	0.430638	0.161895
V4	0.758378	-0.22717
V8	0.673347	-0.03941
V9	0.609152	0.215493
V10	0.568085	-0.32252
V11	0.781738	-0.26073
V12	0.781413	0.277156

Tabela 29 - Cargas fatoriais do poluente MP_{2,5}, período quente e úmido do ano de 2019.

	Fator 1	Fator 2
V1	0.576471	0.122379
V2	0.821011	-0.23413
V3	0.59039	-0.49304
V4	0.804453	0.276207
V5	0.764548	-0.30303
V7	0.869724	0.218939
V8	0.683195	0.221022
V9	0.312025	0.219311
V10	0.6435	0.104375
V11	0.545579	-0.00956
V12	0.708415	-0.06995