

AVALIAÇÃO ESTATÍSTICA MULTIVARIADA APLICADA A QUALIDADE DO AR EM SALVADOR-BA: O CASO DA ESTAÇÃO ITAIGARA

Taiane Campos Braga (*), Édler Lins de Albuquerque

* Instituto Federal da Bahia – IFBA Campus Salvador, taianecbraga@gmail.com

RESUMO

O conhecimento dos valores de concentração dos poluentes é necessário para se concluir que a qualidade do ar de uma determinada região está adequada aos padrões de segurança. Tais informações podem ser fornecidas através da implementação de uma rede de monitoramento da qualidade do ar. Essas concentrações são diretamente afetadas por parâmetros meteorológicos que regem os fenômenos de transporte e difusão de poluentes. Assim, o presente estudo buscou analisar, através de métodos de análise estatística multivariada, a correlação existente entre dados de concentração de poluentes atmosféricos e parâmetros meteorológicos, coletados pela rede de monitoramento da qualidade do ar na cidade de Salvador. A partir disso, os resultados para a avaliação dos períodos seco e chuvoso mostraram que não houve grandes diferenças para as variáveis estudadas, com exceção para o SO₂ que apresentou aumento para o período chuvoso. Além disso, as análises mostraram que as variações anuais para o ozônio no período estudado apresentaram alta entre os meses de julho e outubro em que a radiação solar é menor, indicando a presença de correlação com a velocidade do vento. A análise fatorial por componentes principais, mostrou que houve boas correlações positivas entre as variáveis NO_x, NO₂, NO e CO, além de mostrar a correlação inversa do ozônio, temperatura e velocidade do vento com os óxidos de nitrogênio. Adicionalmente, a análise dos mapas fatoriais mostrou a presença de correlação positiva entre o O₃ com o parâmetro meteorológico velocidade do vento, indicando a ocorrência de transporte do mesmo pelo vento. Em geral duas variáveis não apresentaram correlação com as demais estudadas, sendo estas, o MP₁₀ e o SO₂, indicando a necessidade de maiores estudos com o uso de técnicas que possam avaliar a existência de correlações não lineares para estas. Finalmente, os resultados alcançados com a HCA mostraram concordância com os encontrados pela análise fatorial via PCA reafirmando a importância do uso paralelo de técnicas multivariadas para a melhor compreensão do fenômeno estudado.

PALAVRAS-CHAVE: análise multivariada, qualidade do ar, meteorologia.

INTRODUÇÃO

A qualidade do ar de uma região é avaliada pelo nível de poluição na atmosfera determinado pela quantificação de substâncias consideradas poluentes presentes no ar. Assim, o resultado final da emissão de poluentes, ou seja, o nível de poluição da atmosfera, será diretamente influenciado pelas características das fontes emissoras e pelas interações que essas substâncias sofrerão na atmosfera, sejam essas físicas (diluição) e/ou químicas (reações químicas). Logo, a qualidade do ar pode mudar em razão das condições meteorológicas que determinam uma maior ou menor diluição dos poluentes (CETESB, 2021).

Estudos na área de poluição do ar (LEONI et al. 2016; SANTOS et al, 2018; SONG et al, 2020), confirmam a existência das relações entre as concentrações de poluentes e os principais parâmetros meteorológicos que regem os fenômenos de difusão e diluição. O estudo das interações entre cada parâmetro meteorológico e os dados das concentrações de poluentes no ar podem trazer respostas sobre os diferentes perfis de distribuição e comportamento dos poluentes em diferentes épocas do ano, permitindo a investigação de como padrões atmosféricos podem influenciar as concentrações de poluentes numa dada região tornando-se uma ferramenta integrante essencial para a gestão da poluição do ar local.

Assim, torna-se cada vez mais necessário o uso de técnicas que permitam a facilitação quanto a manipulação de dados meteorológicos para previsão e avaliação do comportamento das concentrações de poluentes emitidos para atmosfera. Um dos recursos que vêm sendo utilizados para o entendimento de tais correlações, têm sido os métodos de análise estatística multivariada que permitem a construção de grupos de elementos amostrais que apresentam similaridade entre si, além de, auxiliar na investigação das relações de dependência entre variáveis respostas associadas ao fenômeno e outros fatores que contribuem na predição de comportamentos existentes nos dados.

A cidade de Salvador possui uma rede de monitoramento da qualidade do ar entre os anos de 2011 e 2016, com oito estações no total. Os parâmetros monitorados por estas estações foram: velocidade e direção do vento, temperatura, umidade relativa do ar, desvio padrão da direção do vento, precipitação pluviométrica (chuva), além das concentrações de dióxido de enxofre (SO₂), monóxido de carbono (CO), ozônio (O₃), material particulado menor que 10 µm (MP₁₀) e óxidos de nitrogênio (NO_x). O presente estudo dispõe-se a analisar as correlações entre os principais parâmetros meteorológicos e as concentrações de poluentes atmosféricos monitorados na cidade de Salvador.

OBJETIVOS

O presente trabalho tem como objetivo analisar, através de métodos de análise estatística multivariada, a correlação existente entre dados de concentração de poluentes atmosféricos e parâmetros meteorológicos, coletados pela rede de monitoramento da qualidade do ar na cidade de Salvador, buscando avaliar a contribuição relativa de variáveis meteorológicas para as alterações nas concentrações de poluentes e; investigar como padrões atmosféricos podem influenciar tais concentrações na cidade de Salvador.

METODOLOGIA

A base para o presente estudo foram os dados gerados pela rede de monitoramento da qualidade do ar presente em Salvador entre os anos de 2011 e 2016. A rede sotropolitana de monitoramento da qualidade ar possuiu ao todo oito estações de monitoramento localizadas nas principais avenidas da cidade de Salvador. A partir disto, foi possível a obtenção de um banco de dados que compreendeu o período de monitoramento entre 01 de janeiro de 2011 a 31 de dezembro de 2016. Tais dados, foram cedidos ao Grupo de Pesquisa em Ecologia Industrial (ECOIN) do Instituto Federal da Bahia pela CETREL S. A. S.A., empresa que operou as estações no período indicado.

Os dados gerados correspondem ao monitoramento horário para treze variáveis (em cada estação), sendo, seis destas, correspondentes aos parâmetros meteorológicos referentes a velocidade do vento (m/s), direção do vento (graus), temperatura (°C), umidade relativa do ar (%), desvio padrão da direção do vento (graus) e precipitação pluviométrica (chuva) (mm), e as outras sete correspondentes aos poluentes atmosféricos dióxido de enxofre – SO₂ (ppb), monóxido de carbono – CO (ppm), ozônio – O₃ (ppb), material particulado menor que 10 µm – MP₁₀ (µg/m³), dióxido de nitrogênio – NO₂ (ppb) óxido nítrico – NO (ppb) e óxidos de nitrogênio totais – NO_x (ppb). Assim, em sua totalidade, o banco de dados bruto gerado abrange 101 variáveis monitoradas possuindo um conjunto amostral de 52.609 observações.

Devido a especificidade da localização de cada estação de monitoramento, optou-se por estudar a correlação entre as variáveis separadamente para cada estação de monitoramento, sendo selecionada para o presente trabalho a estação do Itaigara. A mesma, por sua vez, possuía uma localização estratégica situada na área de ligação entre as Avenidas Antônio Carlos Magalhães e Juracy Magalhães Júnior, onde hoje é encontrado o complexo de viadutos que passaram a fazer a ligação direta entre as vias.

A região escolhida para o estudo possui um intenso tráfego de veículos e está localizada numa área central da cidade, sendo caracterizada como uma área comercial e residencial, onde se encontram estabelecimentos como: shoppings, prédios comerciais, supermercados e hospitais, além de ligar bairros populosos, como Pituba e Itaigara à área da rodoviária, ao aeroporto e à BR-324, possuindo uma alta densidade populacional e uma intensa movimentação de pessoas. Este último, torna-se ainda mais importante quanto a escolha da estação de estudo, já que, níveis altos de poluição podem afetar consideravelmente a saúde da população residente e/ou que circule no local com frequência.

Deste modo, por contar com tais características a estação do Itaigara se mostrou uma das principais estações de monitoramento da qualidade do ar na cidade de Salvador. Através da mesma, foi possível verificar, a partir do monitoramento, como a dispersão de poluentes do ar em áreas urbanas pode ser influenciada por eventos e/ou padrões da meteorologia local.

Para a estação Itaigara, o conjunto de dados gerado diz respeito às médias horárias de doze variáveis, já citadas anteriormente, do período referente a 10 de outubro de 2013 a 03 de março de 2016, compreendendo a um tamanho amostral de 16.597 observações válidas, ou seja, sem a presença de campos vazios ou que não possuíam valores do tipo número. Os dados foram submetidos a tratamento estatístico com o objetivo de se obter medidas características que indicassem o comportamento da distribuição do conjunto de dados. Assim, primeiramente foram utilizadas as medidas de posição ou localização como: medidas de tendência central e separatrizes, com o objetivo de avaliar as medidas de centro da distribuição de valores, além de, medidas de dispersão e variabilidade (erro padrão, desvio padrão, variância) com o intuito de se avaliar o espalhamento dos dados em torno da média.

Todos os procedimentos estatísticos foram aplicados tanto ao banco de dados completo e tratado, que contempla todo o período do monitoramento para a estação Itaigara, como a dois períodos separados caracterizados como seco (primavera e verão) e chuvoso (outono e inverno) de acordo com as normais climatológicas (1981-2010) (INMET, 2010) para o município de Salvador. Tal escolha foi realizada para que fosse possível a avaliação do comportamento de variação dos poluentes atmosféricos com relação à chuva, variável não monitorada na estação escolhida. Logo, para validação deste procedimento os parâmetros meteorológicos para o período monitorado foram então comparados com as normais climatológicas fornecidas pela rede do INMET.

Com o intuito de avaliar a presença de *outliers* nos dados, foram elaborados diagramas de caixa ou Boxplot que indicam as seguintes informações: 1º quartil Q_1 (percentil 25%), 3º quartil Q_3 (percentil 75%), mediana (percentil 50%) e amplitude interquartil (AIQ), além do intervalo considerado *não-outlier* dado pela equação 1.

$$Q_1 - 1,5.AIQ \leq \text{valor} \leq Q_3 + 1,5.AIQ \quad \text{equação (1)}$$

Para valores considerados como outliers as seguintes premissas foram adotadas:

- Se o valor estiver compreendido nas faixas expressas nas equações 2 e 3 o mesmo será considerado um outlier superior ou inferior *moderado*;

$$Q_3 + 1,5.AIQ < \text{valor} < Q_3 + 3.AIQ \quad \text{equação (2)}$$

$$Q_1 - 3.AIQ < \text{valor} < Q_1 - 1,5.AIQ \quad \text{equação (3)}$$

- Se o valor estiver compreendido de acordo com as equações 4 e 5 o mesmo será considerado um outlier superior ou inferior *extremo* e, portanto, retirado das análises estatísticas multivariadas.

$$\text{valor} > Q_3 + 3.AIQ \quad \text{equação (4)}$$

$$\text{valor} < Q_1 - 3.AIQ \quad \text{equação (5)}$$

Após o tratamento estatístico dos dados, os mesmos foram submetidos as técnicas estatísticas multivariadas. Para o presente estudo, duas técnicas estatísticas multivariadas foram aplicadas em conjunto: a análise fatorial por componentes principais (PCA) e a análise hierárquica de agrupamentos (HCA) que são consideradas técnicas complementares, onde informações semelhantes a respeito do conjunto de dados podem ser obtidas por ambas.

Inicialmente, o conjunto de dados foi submetido a padronização via o procedimento *Z-Scores*, onde o valor da variável para uma dada amostra é subtraído da sua média e depois dividido pelo desvio padrão do conjunto amostral. Esse procedimento foi realizado para que houvesse a eliminação da interferência da ordem de grandeza entre as variáveis que apresentam unidades de medidas distintas. Tal procedimento é de essencial importância principalmente quando da aplicação da técnica HCA, já que, a mesma faz uso do cálculo de distâncias, que por sua vez são influenciadas pela ordem de grandeza das variáveis.

A primeira técnica aplicada foi a análise fatorial por componentes principais visando a identificação de correlações entre variáveis originais através da criação de fatores que representam a combinação destas variáveis, promovendo então uma redução ou simplificação estrutural no conjunto de dados originais. Tais fatores foram obtidos por meio da matriz de correlação das variáveis padronizadas e podem ser expressos de acordo com a equação 6, onde k é o número de variáveis e s são os *scores* fatoriais.

$$F_{ki} = s_{1k}.X_{1i} + s_{2k}.X_{2i} + \dots + s_{kk}.X_{ki} \quad \text{equação (6)}$$

Os *scores* fatoriais de cada fator correspondem aos parâmetros de um modelo de regressão linear múltipla, sendo a variável dependente o próprio fator e possuindo como variáveis explicativas, as variáveis padronizadas. O cálculo dos *scores* fatoriais se dá por meio da determinação dos autovalores e autovetores da matriz de correlação dos dados.

Os fatores obtidos a partir da matriz de correlação das variáveis originais padronizadas representam novas variáveis que são combinações lineares das variáveis que as originaram, sendo ortogonais entre si. O critério para escolha do número de fatores escolhidos segue o critério da variância total. No mesmo, é analisada a proporção da variância total relacionada a cada autovalor, permanecendo aqueles autovalores que representarem maiores proporções da variância total do conjunto de dados originais. Dessa forma, os primeiros fatores explicam a maior parcela da variância dos dados, sendo o valor mínimo considerado aceitável maior que 75%. Por fim, após se encontrar os novos fatores, promoveu-se a rotação *varimax* desses, de modo a aproximar os pontos correspondentes as variáveis de um dos novos fatores chamados de fatores rotacionados. Essa rotação, foi realizada com o objetivo de se alcançar uma estrutura mais simples de ser interpretada preservando a orientação original entre os fatores, mantendo-os perpendiculares após a rotação, além de aumentar as correlações entre as variáveis originais e os novos fatores gerados.

A segunda técnica aplicada ao conjunto de dados foi a análise hierárquica de agrupamentos objetivando a divisão dos elementos do conjunto de dados, de forma que os elementos pertencentes a um mesmo grupo sejam similares entre si com respeito às variáveis. Esta técnica visa a alocação das observações em uma quantidade relativamente pequena de

agrupamentos homogêneos internamente e heterogêneos entre si e que representam o comportamento conjunto das variáveis a partir de determinadas observações.

A aplicação da técnica HCA envolve tanto a escolha de um coeficiente de semelhança ou medida de distância, como a escolha de um processo de aglomeração. As medidas de distância consideram que, se dois indivíduos são similares, eles estão próximos um do outro, ou seja, eles são comuns ao conjunto de variáveis e vice-versa. Para o presente trabalho foi utilizada a medida de distância correlação de Pearson. Já o método de aglomeração escolhido foi o método de Ward. O mesmo se distingue dos demais métodos por empregar um procedimento de análise da variância para avaliar as distâncias entre grupos, sendo considerado um método muito eficiente. Assim, os resultados obtidos a partir das análises realizadas com o conjunto de dados para estação Itaipara serão mostrados na seção a seguir.

RESULTADOS

Como dito na seção anterior, inicialmente os dados passaram por tratamento com o intuito de se eliminar as observações que apresentaram valores considerados *outliers extremos*. Para tanto, foram calculados os valores limites para cada faixa, de cada uma das variáveis existentes no banco de dados, conforme tabela 1.

Tabela 1. Valores limites para outliers moderados e extremos. Fonte: Autor do trabalho.

Parâmetros Estatísticos	SO ₂	CO	O ₃	MP ₁₀	NO	NO _x	NO ₂	VV	TEMP	UMID	STWD
1º Quartil	0,00	0,06	5,00	7,70	6,10	11,30	4,40	1,50	23,50	65,00	19,80
3º Quartil	0,40	0,35	11,00	22,70	22,50	33,50	11,30	3,90	26,60	78,00	27,80
AIQ	0,40	0,29	6,00	15,00	16,40	22,20	6,90	2,40	3,10	13,00	8,00
Q ₃ +3.AIQ	1,60	1,22	29,00	67,70	71,70	100,10	32,00	11,10	35,90	117,00	51,80
Q ₃ +1,5.AIQ	1,00	0,79	20,00	45,20	47,10	66,80	21,65	7,50	31,25	97,50	39,80
Q ₁ -1,5.AIQ	--	--	--	--	--	--	--	--	18,85	45,50	7,80
Q ₁ -3.AIQ	--	--	--	--	--	--	--	--	14,20	26,00	--

A partir desta análise foi possível notar que as únicas variáveis que apresentaram *outliers extremos* inferiores foram a temperatura e a umidade relativa do ar, sendo esta a razão destes valores não terem sido expressos na tabela acima para as demais variáveis. Quanto aos *outliers extremos* superiores, todas as variáveis apresentaram a presença dos mesmos. Assim, foi realizada a limpeza do conjunto de dados retirando as amostras consideradas como *outliers extremos* superiores ou inferiores. Após tal procedimento seguiu-se a análise descritiva do conjunto de dados.

Assim, os resultados da estatística descritiva para os dados tratados (sem a presença de *outliers extremos*) estudados nos períodos seco e chuvoso podem ser observados nas tabelas 2 e 3, respectivamente. Pode-se observar que para ambos os períodos avaliados os dados meteorológicos de velocidade do vento, temperatura e umidade relativa do ar apresentaram valores para o coeficiente de assimetria próximos a zero, indicando que os mesmos apresentaram uma tendência a distribuição normal. Analisando-se as medidas de tendências central média, mediana e moda, é perceptível que os mesmos possuem valores próximos para essas variáveis, o que reafirma a disposição dos dados a uma distribuição normal. Entretanto, para as demais variáveis os dados não tenderam a uma distribuição normal, visto que, as mesmas apresentaram maiores coeficientes de assimetria indicando uma distribuição assimétrica positiva (à direita). Este fato é ressaltado quando se observa a relação moda < mediana < média que, foi atendida para as demais variáveis. Tal comportamento foi observado tanto no período seco como no período chuvoso.

Adicionalmente, as variáveis velocidade do vento, temperatura e umidade relativa do ar apresentaram certo grau de achatamento (curva platicúrtica) quando comparadas as demais, pois, o coeficiente de curtose apresentou valores negativos, tanto no período seco quanto no período chuvoso, observando-se um maior achatamento para a velocidade do vento no período chuvoso, que apresentou coeficiente de curtose mais negativo. Para as demais variáveis o coeficiente de curtose é positivo, indicando que a curva de distribuição dos dados apresenta um alongamento, sendo caracterizada como leptocúrtica.

Quanto à variabilidade do conjunto de dados para cada variável analisada, os poluentes atmosféricos SO₂, CO e NO apresentaram valores para o coeficiente de variação mais altos no período seco, seguidos do MP₁₀ no período chuvoso,

indicando alta dispersão relativa no conjunto de dados. A alta variabilidade das variáveis citadas pode ser explicada pelo fato de que as mesmas são predominantemente emitidas por fontes móveis, neste caso, fontes veiculares, devido à proximidade da estação de monitoramento estudada com vias de alta movimentação veicular. Como existem variação do fluxo de veículos nas Avenidas ACM e Juracy Magalhães Júnior ao longo do dia, as emissões desses poluentes também sofrem variação, o que explica o comportamento observado para tais variáveis.

Tabela 2. Estatística descritiva para as variáveis estudadas no período seco. Fonte: Autor do trabalho.

Parâmetros Estatísticos	SO ₂	CO	O ₃	MP ₁₀	NO	NO _x
Mínimo	0,000	0,000	0,000	0,000	0,000	0,000
Média	0,188	0,218	8,095	16,441	14,442	22,327
Mediana	0,100	0,180	7,400	14,200	10,700	18,100
Moda	0,000	0,000	5,300	5,800	4,300	10,900
Máximo	1,600	1,210	27,300	67,400	71,600	96,400
Amplitude	1,600	1,210	27,300	67,400	71,600	96,400
Erro padrão	0,002	0,002	0,043	0,107	0,127	0,163
Desvio padrão	0,241	0,205	4,237	10,540	12,518	16,024
Coefficiente de Variação	1,287	0,942	0,523	0,641	0,867	0,718
Curtose	3,251	1,261	0,438	1,315	2,685	1,548
Assimetria	1,722	1,166	0,764	1,111	1,594	1,264
Parâmetros Estatísticos	NO ₂	VV	TEMP	UMID	STWD	
Mínimo	0,000	0,000	18,500	47,000	4,100	
Média	8,333	2,800	25,419	70,747	27,410	
Mediana	7,300	2,700	25,300	71,000	24,000	
Moda	4,700	2,600	25,300	71,000	21,300	
Máximo	31,800	10,200	31,100	93,000	67,900	
Amplitude	31,800	10,200	12,600	46,000	63,800	
Erro padrão	0,053	0,017	0,022	0,088	0,102	
Desvio padrão	5,243	1,656	2,137	8,675	10,082	
Coefficiente de Variação	0,629	0,591	0,084	0,123	0,368	
Curtose	1,416	-0,138	-0,486	-0,590	1,343	
Assimetria	1,104	0,418	-0,018	0,127	1,332	

Tabela 3. Estatística descritiva para as variáveis estudadas no período chuvoso. Fonte: Autor do trabalho.

Parâmetros Estatísticos	SO ₂	CO	O ₃	MP ₁₀	NO	NO _x
Mínimo	0,000	0,000	0,000	0,400	0,000	0,500
Média	0,329	0,246	9,033	16,017	17,449	25,158
Mediana	0,100	0,200	8,600	12,850	13,400	20,800
Moda	0,000	0,000	7,900	5,400	7,900	9,700
Máximo	1,600	1,220	27,500	67,700	71,700	95,600
Amplitude	1,600	1,220	27,500	67,300	71,700	95,100
Erro padrão	0,005	0,003	0,056	0,152	0,190	0,232
Desvio padrão	0,422	0,217	4,328	11,646	14,606	17,812
Coefficiente de Variação	1,284	0,885	0,479	0,727	0,837	0,708
Curtose	0,920	2,149	0,270	2,184	1,326	0,863
Assimetria	1,385	1,367	0,659	1,411	1,308	1,087
Parâmetros Estatísticos	NO ₂	VV	TEMP	UMID	STWD	

Mínimo	0,000	0,000	18,800	43,000	0,000
Média	8,255	2,701	24,612	72,261	22,034
Mediana	7,400	2,700	24,500	72,000	21,300
Moda	6,600	2,700	25,900	71,000	20,800
Máximo	31,800	8,000	31,100	93,000	66,200
Amplitude	31,800	8,000	12,300	50,000	66,200
Erro padrão	0,068	0,019	0,029	0,120	0,091
Desvio padrão	5,205	1,454	2,211	9,239	7,006
Coefficiente de Variação	0,631	0,538	0,090	0,128	0,318
Curtose	1,144	-0,425	-0,386	-0,552	5,338
Assimetria	0,995	0,166	0,242	0,094	1,434

Após o estudo do conjunto de dados a partir da estatística descritiva, as técnicas multivariadas foram enfim aplicadas ao conjunto de dados. A primeira técnica a ser aplicada foi a análise fatorial via componentes principais, onde, para as dez variáveis consideradas na análise realizada, foi possível a extração de quatro fatores, os quais explicaram aproximadamente cerca de 76,7% da variância total dos dados. Os autovalores e o resumo das porcentagens da variância explicada pelos fatores extraídos são apresentados na tabela 4.

Tabela 4. Autovalores e explicação da variância total. Fonte: Autor do trabalho.

Componentes	Autovalor	Variância (%)	Variância Acumulada
1	3,541063	35,41063	35,4106
2	2,061039	20,61039	56,0210
3	1,095680	10,95680	66,9778
4	0,974995	9,74995	76,7278
5	0,755014	7,55014	84,2779
6	0,612246	6,12246	90,4004
7	0,465831	4,65831	95,0587
8	0,339432	3,39432	98,4530
9	0,151768	1,51768	99,9707
10	0,002932	0,02932	100,0000

Quando testada a possibilidade de extração de maior número de componentes, aumentos progressivos não mostraram resultados mais esclarecedores, mostrando ser adequada a extração e apenas quatro. A consideração de um maior número de fatores, mesmo com a utilização do recurso de rotação de fatores pela metodologia *varimax*, mostrou que os resultados não apresentaram maiores benefícios para interpretação, visto que, a extração de componentes que expliquem apenas uma única variável não apresenta nenhum benefício quanto ao entendimento das correlações para o fenômeno estudado.

A partir da extração dos fatores, foram então analisados os valores para os coeficientes de cada fator não rotacionado, mostrados na tabela 5. Como pode ser observado, no primeiro fator é evidenciada a formação de dois grupos de variáveis que possuem os maiores coeficientes, sendo NO_x, NO₂, NO seguidos do CO e ozônio que apresentou coeficiente negativo. O segundo fator apresentou os maiores coeficientes para as variáveis velocidade do vento, temperatura e umidade relativa do ar. Já os dois últimos fatores apresentaram coeficientes altos apenas para o MP₁₀, no terceiro fator e, SO₂, no quarto fator.

Tabela 5. Resultados para os coeficientes dos fatores principais não rotacionadas. Fonte: Autor do trabalho.

	Factor 1	Factor 2	Factor 3	Factor 4
SO ₂	0,169388	-0,114566	0,386183	0,881285
CO	0,604738	-0,310115	0,073998	-0,164267
O ₃	-0,682672	-0,142510	0,509312	-0,068700

MP	0,182466	-0,327576	0,629330	-0,387857
NO	0,872690	-0,269840	-0,112998	0,055519
NOx	0,919776	-0,291422	-0,026968	0,021343
NO ₂	0,733187	-0,262288	0,170648	-0,075150
Vel. Vento	-0,494647	-0,629631	0,149740	-0,005849
Temp.	-0,308205	-0,742620	-0,447832	0,040722
Umidade	0,403140	0,806023	0,141495	-0,072915

A partir disso, foram elaborados os mapas fatoriais para os quatro primeiros fatores não rotacionados. Nos mapas fatoriais, as retas mostradas indicam os ângulos formados pela posição de cada variável em relação à origem dos eixos das componentes principais. Quanto menor for o ângulo formado pelas retas para um par de variáveis, maior é a semelhança entre o comportamento dessas variáveis, mostrando uma maior correlação linear. Em contraste, ângulos próximos de 180° demonstram uma correlação linear inversa, tendo as variáveis investigadas coeficientes de correlação negativos (OTTO, 2017).

A figura 1, traz o mapa fatorial para dois primeiros fatores extraídos evidenciando que, as variáveis NO_x, NO₂, NO e CO estão positivamente correlacionadas. A correlação apresentada pelas variáveis é explicada pelo fato destes poluentes serem emitidos pelas mesmas fontes, que no caso da região estudada são as fontes veiculares. Ao se analisar as estatísticas descritivas nota-se que, tais variáveis, apresentaram picos de concentração nos horários em que o fluxo de veículos é maior, como pode ser visto na figura 2, o que afirma a existência da contribuição de uma mesma fonte na emissão desses poluentes.

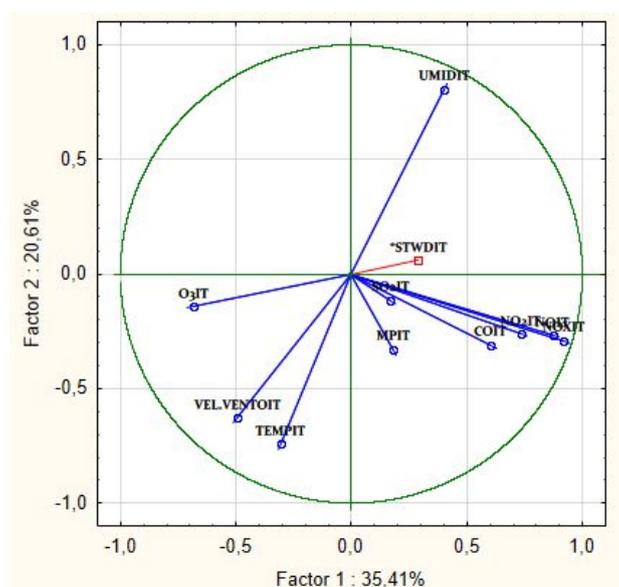
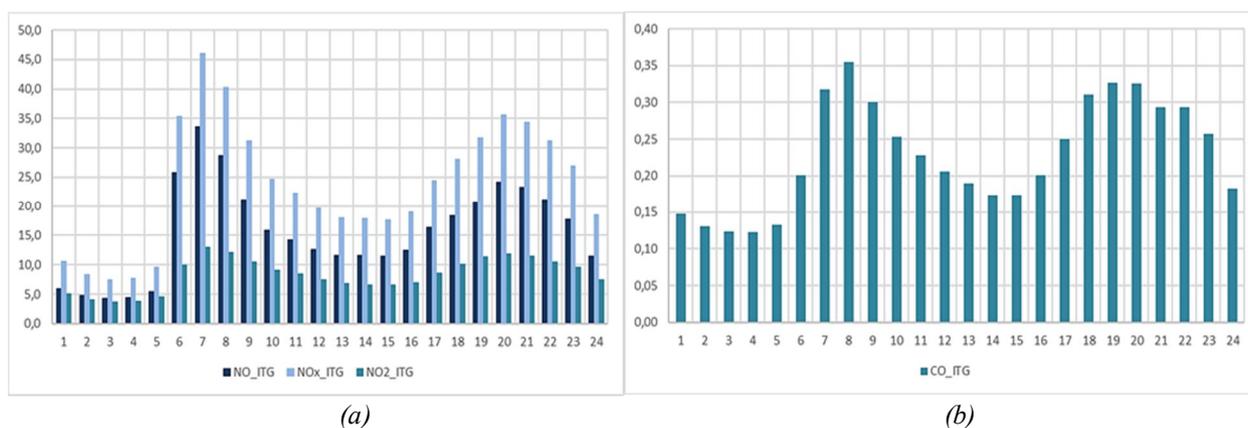


Figura 1. Mapa fatorial para o Fator 1 e Fator 2. Fonte: Autor do trabalho.



(a)

(b)

Figura 2. Variação horária para óxidos de nitrogênio (a) e CO (b) na estação Itaigara. Fonte: Autor do trabalho.

O mapa para o primeiro e terceiro fator, figura 3, mostra correlação positiva para a velocidade do vento e o ozônio, indicando a ocorrência de transporte de ozônio, sendo uma parte do mesmo proveniente do arraste provocado pelo vento, originado de localidades vizinhas a região do Itaigara. É possível notar também que o ozônio apresenta uma correlação negativa com os óxidos de nitrogênio, ou seja, aumentos nas concentrações de óxidos de nitrogênio acarretam diminuição na concentração de ozônio. Tal correlação está ligada ao fato de que o ozônio é consumido em reações químicas, sendo a principal delas conhecida como *ozone titration*, (AHRENS & HENSON, 2019).

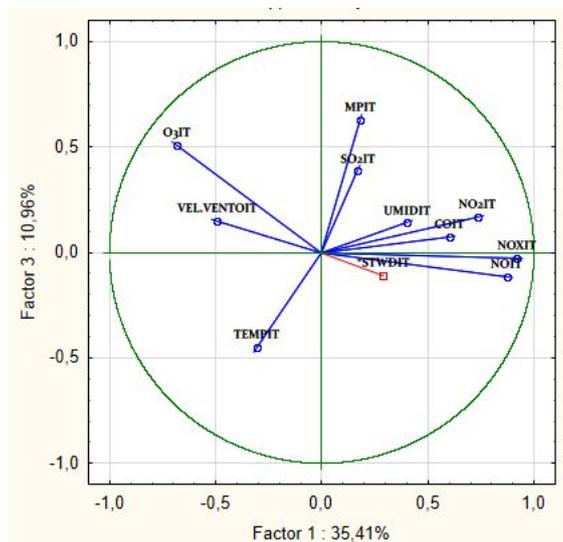


Figura 3. Mapa fatorial para o Fator 1 e Fator 3. Fonte: Autor do trabalho.

A correlação das concentrações de ozônio com a variável meteorológica velocidade do vento é melhor entendida quando observadas a variação anual para estas duas variáveis, como pode ser visto na figura 4.

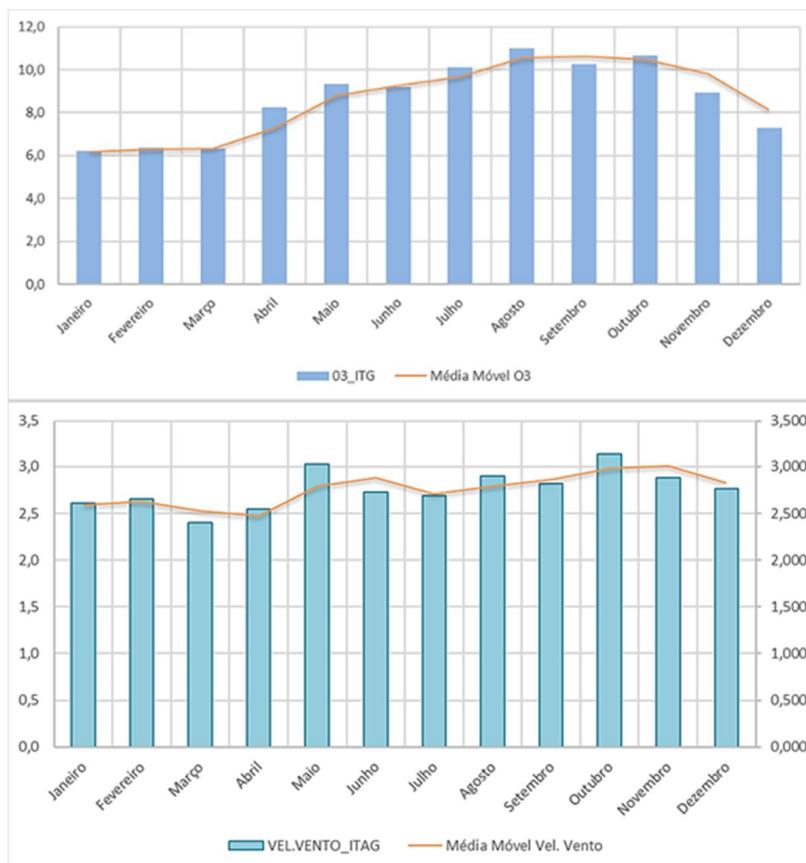


Figura 4. Variação anual das concentrações de O₃ (ppb) e velocidade do vento (m/s) para a estação Itaigara. Fonte: Autor do trabalho.

Os meses de junho a agosto são meses em que a radiação solar apresenta níveis menores comparados a outros períodos, indicando que as concentrações de O_3 encontradas na região estudada podem estar correlacionadas com outros parâmetros meteorológicos, neste caso, a velocidade do vento.

A formação de ozônio está diretamente relacionada com a intensidade de radiação solar e concentração de seus precursores, esperando-se concentrações maiores nos períodos de maior radiação solar, o que difere dos resultados encontrados. Sendo assim, ao se analisar dados de velocidade (m/s) e direção (graus) para o vento na região, verifica-se que no período em que foram encontradas maiores concentrações para o ozônio, foi verificado a presença de classes de vento de maiores velocidades, conforme figura 5, além do fato do perfil das concentrações de O_3 seguirem o perfil de direção nos períodos tanto de alta no valor das concentrações como de baixa, conforme mostrados pelas figura 6, indicando que o ozônio encontrado na região pode estar sendo transportado de localidades vizinhas, tendo uma parte de sua produção em regiões de maior concentração de NO_x e compostos orgânicos, que são seus precursores.

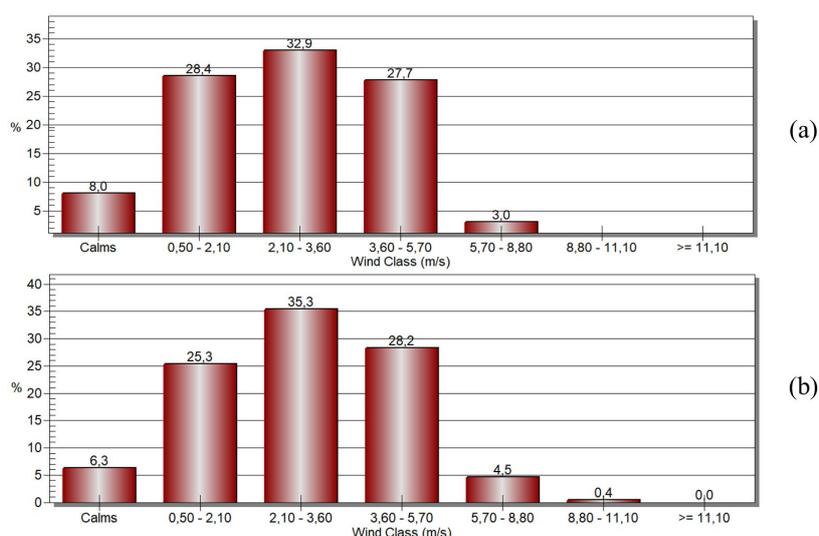


Figura 5. Distribuição de frequência para as classes de velocidade do vento - (a) menores concentrações de O_3 (nov/jun) | (b) maiores concentrações de O_3 (jul/out). Fonte: Autor do trabalho.

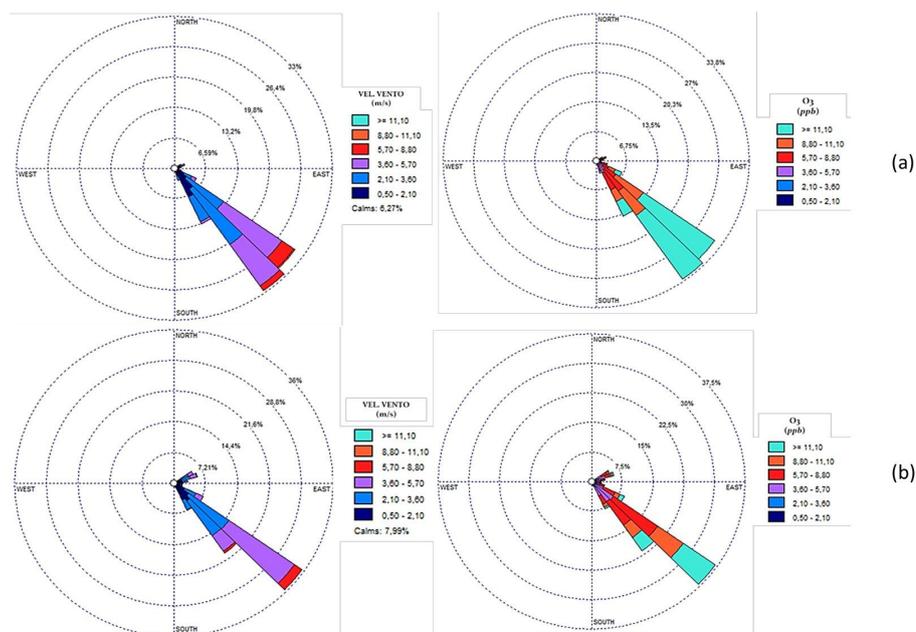


Figura 6. Rosa dos ventos e variação da concentração de O_3 – (a) novembro a junho | (b) julho a outubro. Fonte: Autor do trabalho.

Em geral duas variáveis não apresentaram correlação com as demais estudadas, sendo estas, o MP_{10} e o SO_2 . O mapa fatorial para o primeiro e quarto fator, na figura 7, apresenta o distanciamento dessas variáveis em relação às outras, fato que pode ser visto pela formação de ângulos próximos a 90° graus para o MP_{10} e o SO_2 com as demais retas, mostrando que as análises realizadas até o momento não se mostraram conclusivas para as mesmas. É possível observar ainda, a

correlação negativa entre a maioria dos poluentes e velocidade do vento, temperatura e ozônio. Isso se explica pelo fato de que a maioria destes poluentes, no caso os óxidos de nitrogênio, fazem parte de reações fotoquímicas sendo precursores do ozônio, tendo suas concentrações diminuídas quando consumidos nestas reações. Outro fator é que maiores velocidades do vento contribuem para uma maior dispersão dos mesmos na atmosfera o que também gera a diminuição nas concentrações, explicando a correlação negativa encontrada.

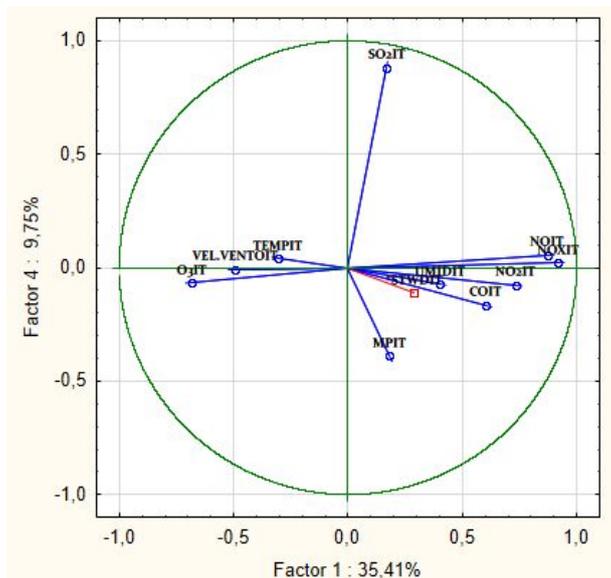


Figura 7. Mapa fatorial para o Fator 1 e Fator 4. Fonte: Autor do trabalho.

Por fim, aplicou-se ao mesmo conjunto de dados padronizados, a análise hierárquica de agrupamentos (HCA) com o objetivo de reforçar os resultados para a PCA além de trazer novas formas de interpretação dos dados. Assim, adotando-se a medida de distância 1 – Pearson-r e o critério de Ward foi possível a alocação das variáveis em clusters (agrupamentos) homogêneos internamente e heterogêneos entre si. Na figura 8, é apresentado o dendrograma para as variáveis estudadas ilustrando os grupos formados. No mesmo, é possível a visualização de três grandes agrupamentos ao considerar a altura de corte em 40.

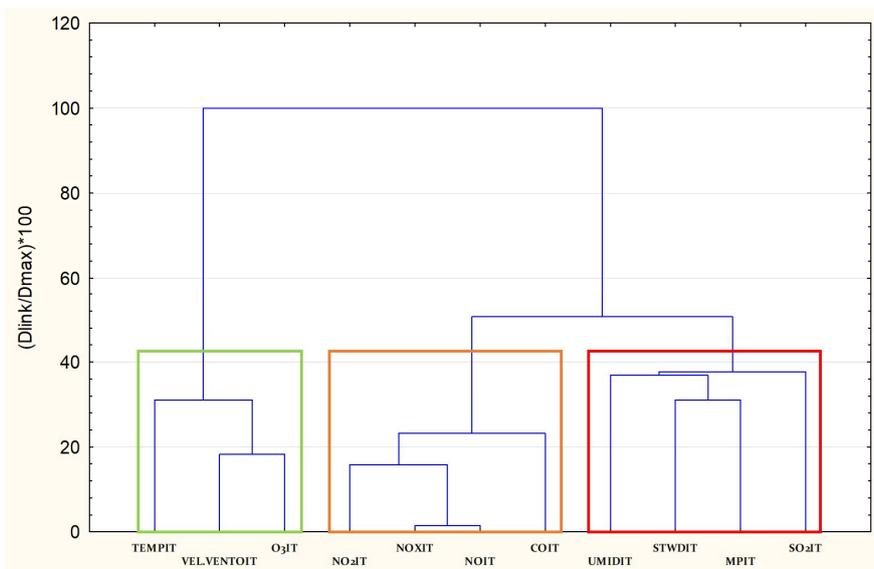


Figura 8. Dendrograma com os clusters para as variáveis estudadas (1 – Pearson-r). Fonte: Autor do trabalho.

No primeiro grupo formado à direita (em vermelho), nota-se o agrupamento das variáveis MP₁₀ e STWD, seguido do agrupamento das mesmas com a umidade relativa do ar e o SO₂. Maiores valores de STWD podem indicar maior instabilidade na atmosfera da região, favorecendo a dispersão de poluentes. Além disso, tanto o MP₁₀ e SO₂, podem sofrer alterações em suas concentrações devido a grandes variações de umidade relativa do ar e precipitação. No caso do material particulado, o mesmo pode ser removido da atmosfera por lavagem em dias chuvosos (umidade relativa do ar relativa 100%) ou por deposição úmida. Já o SO₂ pode ser oxidado a sulfato formando aerossol, além de ser um composto bastante

solúvel em água sofrendo diminuição em sua concentração devido a remoção úmida. Entretanto, analisando-se os resultados obtidos anteriormente, para a análise fatorial por componentes principais, torna-se necessário ainda a aplicação de outras técnicas que possam avaliar tais correlações que podem ser melhor descritas por métodos não lineares.

O outro grupo formado, destacado em laranja, corresponde aos óxidos de nitrogênio e CO, apresentando a mesma correlação encontrada na análise fatorial via PCA. Tal resultado reafirma a contribuição das fontes veiculares locais, já que, na localidade estudada não existem fontes fixas. O último cluster formado agrupa as variáveis O₃ e velocidade do vento, seguidos da temperatura. O resultado encontrado reforça que o ozônio, poluente secundário, possui uma parcela de sua concentração tendo como origem o transporte pelos ventos.

Por fim, de modo geral, verificou-se tendências semelhantes entre as análises multivariadas empregadas, mostrando que as mesmas serviram como ferramenta para confirmar algumas tendências observadas anteriormente com relação as concentrações de alguns poluentes trazendo novos recursos de interpretação dos resultados.

CONCLUSÕES

A análise fatorial por componentes principais, mostrou que houve correlações positivas entre as variáveis NO_x, NO₂, NO e CO, o que pode ser entendido como contribuição significativa de uma mesma fonte de emissão. Além disso, a PCA mostrou a presença de correlação inversa do ozônio, temperatura e velocidade do vento com os óxidos de nitrogênio estudados, estando a mesma ligada ao consumo do ozônio em reações químicas com estes compostos e ao fato de que maiores velocidades de vento contribuem para a dispersão desses poluentes. A análise dos mapas fatoriais mostrou a presença de correlação positiva entre o O₃ com o parâmetro meteorológico velocidade do vento, indicando a ocorrência de transporte do mesmo pelo vento, sendo tal resultado validado pela análise da variação anual das concentrações de ozônio e velocidade do vento.

Em geral duas variáveis não apresentaram correlação com as demais estudadas, sendo estas, o MP₁₀ e o SO₂, mostrando a necessidade de maiores estudos que façam uso de técnicas mais robustas que possam avaliar a existência de correlações não lineares entre estas e as demais variáveis estudadas, como é o caso da análise de correlações canônicas.

Os resultados alcançados com a HCA mostraram concordância com os encontrados pela análise fatorial via PCA, sendo obtidos com a mesma quatro grandes grupos ou clusters principais correspondentes a: óxidos de nitrogênio e CO, reafirmando a contribuições de uma mesma fonte poluidora; O₃ e velocidade do vento, reiterando a contribuição do vento para o transporte de ozônio; e o último grupo formado por MP₁₀ e SO₂ e umidade relativa do ar relativa, diferindo das demais análises anteriores. A aplicação da HCA reiterou ainda a necessidade do uso de outras técnicas que possam explicar de forma mais clara o comportamento par as variáveis MP₁₀ e SO₂.

Adicionalmente, foi verificado que a partir dos resultados para as concentrações dos poluentes estudados, foi possível verificar a influência das fontes de emissão nos valores de concentração avaliados, mostrando a importância da presença de dados de fluxo de veículos para a complementação das análises realizadas. Assim, estudos que possam avaliar a correlação da variação nas taxas de emissão relacionadas às fontes poluidoras podem trazer maiores esclarecimentos quanto aos padrões de comportamento para as variáveis estudadas.

A avaliação quanto a aplicação de técnicas estatísticas multivariadas, assim como das análises estatísticas descritivas para a análise de dados de monitoramento da qualidade do ar, revela que o uso de tais técnicas representa uma ferramenta adicional de gestão ambiental, visto que as mesmas fornecem informações sobre a qualidade do ar local, mostrando como a combinação dos fatores fontes de emissão, poluentes e fenômenos atmosféricos podem contribuir para o cumprimento, ou não, dos padrões ambientais que asseguram a segurança da população e meio ambiente de uma dada região. Dessa forma, tais estudos podem auxiliar tanto na avaliação sobre a operação de uma determinada rede de monitoramento já instalada, assim como pode contribuir quanto a decisão da alocação de novas estações de monitoramento e/ou mudanças na disposição dessas estações constituintes da rede de monitoramento da qualidade do ar atual.

Outro fator importante de ser observado sobre as técnicas utilizadas é que as mesmas podem ainda contribuir para a interpretação de resultados de estudos de dispersão de poluentes atmosféricos, que por sua vez se mostram como alternativa quanto a avaliação da qualidade do ar local frente a instalação de uma rede de monitoramento, onde tais técnicas trazem informações sobre como as concentrações dos poluentes podem ser afetadas por parâmetros meteorológicos que regem os fenômenos atmosféricos locais. Adicionalmente as mesmas podem ainda ser utilizadas para o tratamento de dados, tais como, detecção de outliers, tratamento de dados históricos e identificação de padrões nos mesmos.

REFERÊNCIAS BIBLIOGRÁFICAS

1. AHRENS, C. D.; HENSON. **Meteorology Today: An Introduction to Weather, Climate, and the Environment**. 12^a. ed. Boston: Cengage, v. único, 2019.
2. Companhia Ambiental do Estado de São Paulo - CETESB. **Qualidade do ar no Estado de São Paulo 2020**. CETESB: 2021.
3. FÁVERO, L. P. L.; BELFIORE, P. P. **Manual de análise de dados: estatística e modelagem multivariada com Excel, SPSS e Stata**. 1^a. ed. Rio de Janeiro: Elsevier, v. único, 2017.
4. HAIR, J. F. **Análise multivariada de dados**. 6^a. ed. Porto Alegre: Bookman, v. único, 2009.
5. INMET. **Normais Climatológicas**. INSTITUTO NACIONAL DE METEOROLOGIA, 2010. Disponível em: <https://clima.inmet.gov.br/GraficosClimatologicos/BA/83377>>. Acesso em: 17 nov. 2020.
6. LEONI, C.; SAMPAIO, N. A. D. S.; CORRÊA. **Estatística Multivariada Aplicada ao Estudo da Qualidade do Ar**. Revista Brasileira de Meteorologia, Rio de Janeiro, v. 32, n. 2, p. 235-241, jun. 2017.
7. MINGOTI, S. A. **Análise de Dados Através de Métodos de Estatística Multivariada: Uma Abordagem Aplicada**. 1^a. ed. Belo Horizonte: UFMG, v. único, 2005.
8. OTTO. **Chemometrics: Statistics and Computer Application in Analytical Chemistry**. 3^a. ed. Weinheim: Wiley-VCH Verlag GmbH & Co.KGaA, v. único, 2017.
9. SANTOS, F. S. D. et al. **Avaliação da influência das condições meteorológicas na concentração de material particulado fino (MP2,5) em Belo Horizonte, MG**. Engenharia Sanitária e Ambiental, Belo Horizonte, v. 24, n. 2, p. 371-381, jun. 2018.
10. SONG; DENG; REN, Z. **Correlation and principal component regression analysis for studying air quality and meteorological elements in Wuhan, China**. Environ Prog Sustainable Energy, Wuhan, v. 39, n. 1, p. e13278, fevereiro, 2020.