

Artigo

## Estatística Multivariada Aplicada ao Estudo da Qualidade do Ar

Roberto Campos Leoni<sup>1</sup>, Nilo Antonio de Souza Sampaio<sup>2</sup>, Sergio Machado Corrêa<sup>2</sup><sup>1</sup>*Academia Militar das Agulhas Negras, Resende, RJ, Brasil.*<sup>2</sup>*Faculdade de Tecnologia, Universidade do Estado do Rio de Janeiro, Resende, RJ, Brasil.*

Recebido em 5 de Março de 2016 – Aceito em 17 de Novembro de 2016

### Resumo

O estudo da qualidade do ar de uma cidade envolve diferentes aspectos, como as emissões, as transformações físico-químicas, a meteorologia e a topografia, tornando esse estudo altamente não linear. Em particular, as cidades brasileiras experimentam um mosaico bem distinto do panorama mundial, em função do uso de diferentes combustíveis por nossa frota veicular. Neste estudo, utilizou-se um conjunto de dados coletados por uma estação de monitoramento da qualidade do ar na cidade do Rio de Janeiro, no bairro da Gávea, entre julho e outubro de 2011, em um total de 2240 observações de médias horárias de 11 variáveis. Os dados foram tratados estatisticamente, de forma descritiva e multivariada, de modo a elucidar a correlação entre as variáveis envolvidas. A estatística multivariada empregou a análise dos componentes principais e agrupamentos euclidianos e critérios de Ward. Os resultados indicaram uma alta correlação entre os poluentes primários óxidos de nitrogênio e monóxido de carbono (0,71), evidenciando que possuem a mesma origem veicular e que o monóxido de nitrogênio e o ozônio (0,66) e este com a radiação solar também estão correlacionados (0,41), corroborando a formação fotoquímica do ozônio. Outras correlações são entre a temperatura e a umidade relativa do ar (0,64) e que o ozônio possui uma contribuição de localidades vizinhas, em função da dependência deste com a velocidade do vento (0,59).

**Palavras-chave:** poluição do ar, análise multivariada, meteorologia, ozônio.

## Multivariate Analysis Applied to Air Quality Study

### Abstract

The air quality study in a city involves different aspects, such as emissions, physical and chemical processes, meteorology and topography, making this study highly nonlinear. In particular, Brazilian cities experience a very different mosaic from world scenario, due to the use of different fuels by our vehicular fleet. In this study, it was used a data set collected by an air quality monitoring station in the city of Rio de Janeiro, at Gávea district, between July and October of 2011, in a total of 2,240 hourly average data set. These data were treated statistically, using a descriptive and multivariate approach in order to elucidate the correlation between the variables involved. Multivariate statistical used the principal component analysis and cluster Euclidean with Ward criteria. The results showed a strong correlation between the primary pollutants nitrogen monoxide and carbon monoxide (0.71), showing that they have the same vehicular origin and nitrogen monoxide and ozone (0.66) and this with solar radiation (0.41), corroborating the photochemical formation of ozone. Other correlation is with temperature and relative humidity (0.64) and that ozone has a contribution of neighboring localities, due to the dependence of this with the wind speed (0.59).

**Keywords:** air pollution, multivariate analysis, meteorology, ozone.

### 1. Introdução

A qualidade do ar de uma grande cidade é o resultado das emissões atmosféricas das fontes antropogênicas fixas e móveis (veicular), das fontes naturais, dos processos de deposição via seca e úmida, do transporte entre localidades, da topografia e das transformações físico-químicas que

ocorrem na atmosfera, com a conversão dos poluentes primários emitidos em poluentes secundários, como é o caso do ozônio (Orlando *et al.*, 2010).

A química da atmosfera é um tema complexo e pode ser abordado de diferentes enfoques. Porém, os resultados são de difícil interpretação, pois a química da atmosfera é

altamente não linear (Teixeira *et al.*, 2012; Martins *et al.*, 2015). Os modelos de qualidade do ar e de previsão podem ser classificados de diferentes modos (Lora, 2002). Pode-se classificar um modelo por sua estrutura básica em determinísticos, estocásticos ou estacionário dependente do tempo. Outra classificação é por seu marco de referência, como um modelo Euleriano ou Lagrangiano. Com relação à sua dimensionalidade, tem-se os modelos adimensionais, unidimensionais, bidimensionais ou tridimensionais. Finalmente, pelo método de resolução das suas equações fundamentais, tem-se os modelos analíticos ou numéricos.

A química da atmosfera urbana tem seu foco nos compostos orgânicos voláteis (COV) e suas reações fotoquímicas envolvendo os óxidos de nitrogênio ( $\text{NO}_x = \text{NO} + \text{NO}_2$ ), com a conseqüente formação do ozônio ( $\text{O}_3$ ) troposférico e uma série de COV oxidados (Finlayson-Pitts e Pitts, 2000). Nestas reações o radical hidroxila ( $\bullet\text{OH}$ ) reage rapidamente com os COV antropogênicos formando radicais intermediários  $\bullet\text{RO}_2$  e  $\bullet\text{HO}_2$ , que reagem com o NO convertendo-o a  $\text{NO}_2$ , que decompõe fotoquimicamente a um átomo de oxigênio excitado  $\text{O}(^3\text{P})$  e NO. A reação do  $\text{O}(^3\text{P})$  com o oxigênio molecular ( $\text{O}_2$ ) é a única fonte antropogênica do ozônio na troposfera (Atkinson, 2000). A decomposição do ozônio forma outro átomo de oxigênio excitado  $\text{O}(^1\text{D})$  que reage com o vapor d'água formando os radicais  $\bullet\text{OH}$  que reiniciam todo o processo.

Neste trabalho pretende-se focar a qualidade do ar não de um prisma de modelos de qualidade do ar e sim com uma abordagem estatística multivariada, onde será analisada a interdependência dos dados coletados em uma estação automática da qualidade do ar na cidade do Rio de Janeiro.

## 2. Metodologia

Os dados foram coletados por uma estação móvel de monitoramento da qualidade do ar da Secretaria Municipal de Meio Ambiente (SMAC) da Prefeitura da Cidade do Rio de Janeiro, localizada na PUC-Rio (Rua Marquês de São Vicente, 225, Gávea, 22°58'44" S e 43°13'54" W, altitude 20 m), entre os meses de julho a outubro de 2011, compreendendo as estações do inverno e primavera. Cabe ressaltar que também está presente nas proximidades da PUC-Rio uma das vias trânsito mais movimentadas da cidade, a Auto Estrada Lagoa – Barra. Segundo o site da Prefeitura do Rio de Janeiro, esta via possui um fluxo diário de aproximadamente 130 mil veículos. De acordo com a localização podem ser esperados massas de ar mais concentradas de poluentes vindos, de forma decrescente, do Jardim Botânico e Botafogo, da Auto Estrada Lagoa Barra, da montanha (Rocinha), do Leblon, do mar e da floresta da Tijuca. Sendo assim, esta região sofre influência de diferentes tipos de fontes de emissão (Luna *et al.*, 2014).

As variáveis consideradas neste estudo foram: dióxido de nitrogênio ( $\text{NO}_2$ ), monóxido de nitrogênio (NO), óxidos de nitrogênio ( $\text{NO}_x$ ), monóxido de carbono (CO),

ozônio ( $\text{O}_3$ ), velocidade escalar do vento (VEV), radiação solar global (RSG), temperatura (TEM), umidade relativa (UR) e precipitação pluviométrica (PP). A variável PP foi utilizada para determinar a retirada dos dados dos dias chuvosos e quando o seu valor se apresentava diferente de zero, estes dados eram expurgados do banco, pois a chuva reduz o nível dos poluentes atmosféricos solúveis em água por deposição úmida. Além disso, os dados obtidos nos sábados, domingos e feriados foram também desconsiderados, pelo reduzido tráfego veicular. Os dados noturnos também foram removidos, pois à noite a camada de mistura da atmosfera é muito baixa, o que aumenta a concentração de alguns poluentes primários e também a reação fotoquímica é praticamente inexistente, produzindo níveis de ozônio muito reduzidos. A atmosfera urbana noturna deve ser um estudo à parte, por suas peculiaridades, com grande atividade do radical nitrato. Resumindo, foram tratados estatisticamente as médias horárias dos dados compreendidos entre 6:30 h e 18:30 h dos dias úteis e com precipitação pluviométrica igual a zero, em um total de 870 observações com 11 variáveis.

Para as medições do CO foi utilizado o analisador Ecotech modelo EC9830, que realiza medidas na faixa de 0 a 200 ppm com limite de detecção de 50 ppb. Para as medições dos  $\text{NO}_x$  foi utilizado o analisador Ecotech modelo EC9841, com medidas na faixa de 0 a 20 ppm com limite de detecção de 0,5 ppb. Para as medições do  $\text{O}_3$  foi utilizado o analisador da Ecotech modelo EC9810 que opera na faixa de 0 a 20 ppm com limite de detecção de 0,5 ppb.

Em Estatística Multivariada, a análise de agrupamentos representa um conjunto de técnicas exploratórias. São ferramentas úteis e que podem ser aplicadas quando há a intenção de se verificar a existência de comportamentos semelhantes entre variáveis. O objetivo de se criarem grupos, ou clusters, é para prevalecer a homogeneidade interna. Nesse sentido, esse conjunto de técnicas, também conhecido por análise de conglomerados ou análise de clusters, tem por objetivo principal a alocação de observações em uma quantidade relativamente pequena de agrupamentos homogêneos internamente e heterogêneos entre si e que representam o comportamento conjunto das observações a partir de determinadas variáveis (Fávero e Belfiore, 2015).

O entendimento dos resultados de um experimento envolve a análise de grande número de variáveis. Muitas vezes, um pequeno número destas variáveis contém as informações relevantes, enquanto que a maioria das variáveis adiciona pouco ou nada à interpretação dos resultados em termos práticos. A decisão sobre quais variáveis são importantes é feita, geralmente, com base na experiência, ou seja, baseado em critérios que são mais subjetivos que objetivos. A redução de variáveis através de critérios objetivos, permitindo a construção de gráficos bidimensionais contendo maior informação estatística, pode ser conseguida através da análise de componentes principais. Também é possível construir agrupamentos entre as amostras de acor-

do com suas similaridades, utilizando todas as variáveis disponíveis, e representá-los de maneira bidimensional através de um dendograma. A análise de componentes principais e de agrupamento hierárquico são técnicas de estatística multivariada complementares que têm grande aceitação na análise de dados experimentais (Neto e Moita, 1997).

Duas técnicas estatísticas multivariadas foram combinadas na condução das análises: métodos de componentes principais e agrupamento hierárquico para descrever e visualizar a semelhança entre as variáveis. As técnicas foram implementadas com o pacote *FactoMineR* (Lê *et al.*, 2008) disponível na Linguagem R (R Core Team, 2013). Após a seleção dos componentes principais, o agrupamento hierárquico foi realizado com base na medida de distância euclidiana e o critério aglomeração de Ward, pois buscou-se gerar grupos (clusters) que possuam uma alta homogeneidade interna.

### 3. Resultados e Discussão

As medidas descritivas média, desvio padrão e coeficiente de variação das variáveis são apresentadas na Tabela 1. Observa-se que o coeficiente de variação da variável CO é o mais alto quando comparado com as demais variáveis, indicando alta dispersão relativa, seguida por RSG e NO. Os dados referem-se a 870 médias horárias dos dados entre Julho e Outubro de 2011, para o período entre 6:30 as 18:30 h, dos dias úteis e sem chuva.

Além da variação ao longo do dia das atividades ao redor da estação de monitoramento, como o trânsito nas vias ao redor e a ocupação do estacionamento onde se localizava a estação de monitoramento, pode-se também explicar a alta variabilidade do CO e do NO por estas serem as espécies inorgânicas predominantemente emitidas pelas fontes móveis, juntamente com os COV, que não foram mensurados no conjunto de dados estudado. A variabilidade da RSG é explicada pela diferença de intensidade do sol ao longo do dia, onde nas primeiras horas da manhã e nas horas finais da tarde são observados períodos de sombra.

**Tabela 1** - Estatística descritiva dos dados.

Variáveis	Média	Desvio Padrão	Coeficiente de Variação
NO <sub>2</sub>	21,89 µg m <sup>-3</sup>	11,27 µg m <sup>-3</sup>	51%
NO	43,03 µg m <sup>-3</sup>	35,16 µg m <sup>-3</sup>	82%
NOx	65,11 µg m <sup>-3</sup>	39,87 µg m <sup>-3</sup>	61%
CO	0,14 ppm	0,19 ppm	136%
O <sub>3</sub>	40,70 µg m <sup>-3</sup>	25,48 µg m <sup>-3</sup>	63%
VEV	1,08 m s <sup>-1</sup>	0,51 m s <sup>-1</sup>	47%
RSG	311,7 W m <sup>-2</sup>	269,8 W m <sup>-2</sup>	87%
TEM	23,25 °C	3,16 °C	14%
UR	69,13%	15,10%	22%

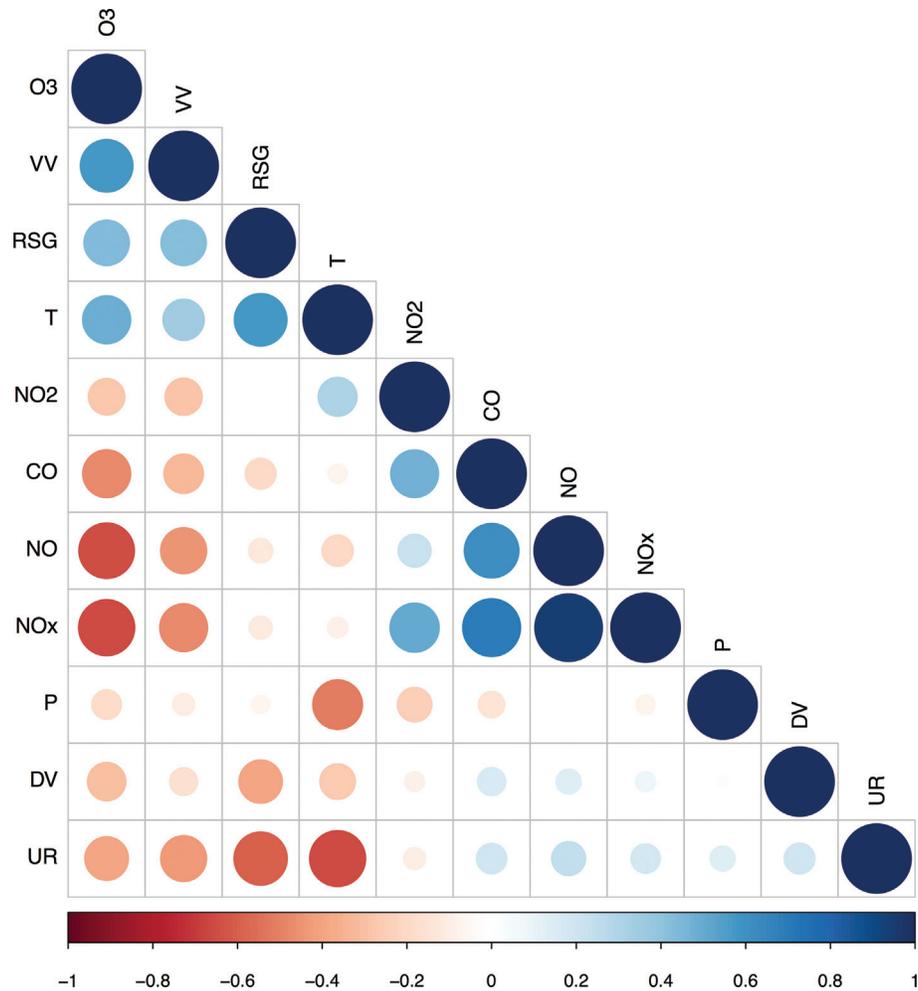
A Fig. 1 ilustra as correlações entre as variáveis. As variáveis NOx e NO apresentam alto grau de correlação linear positiva ( $r = 0,95$ ), justificado pelo fato do NOx ser a soma do NO e NO<sub>2</sub>, ao passo que as variáveis NOx com RSG e TEM praticamente são não correlacionadas. Altas correlações justificam o uso dos componentes principais na redução das dimensões. Entretanto, variáveis não correlacionadas aos pares inviabilizam o uso dos componentes principais, tornando-se inapropriados para redução da dimensionalidade. Conforme discutem Hair *et al.* (2009) e segundo Fávero e Belfiore (2015), embora a inspeção visual da matriz de correlações não revele se a extração de fatores será, de fato, adequada, uma quantidade substancial de valores inferiores a 0,30 representa um preliminar indício de que a análise fatorial poderá ser inapropriada. Para que seja verificada a adequação global propriamente dita da extração dos fatores, deve-se recorrer ao teste de esfericidade de Bartlett (1950). Esse teste foi empregado com a finalidade de avaliar a hipótese nula de esfericidade. A rejeição da hipótese nula de esfericidade indica que é apropriado reduzir a dimensionalidade dos dados. A estatística teste de Bartlett calculada com base nas variáveis conduziu os resultados à rejeição da hipótese nula de esfericidade ( $\chi^2_{36} = 6288,2 p\text{-valor} \cong 0$ ), ou seja, justifica-se o uso de componentes principais para reduzir a dimensão.

Os autovalores e um resumo das porcentagens da variância total explicada pelas componentes principais são apresentados na Tabela 2. Observa-se que os três primeiros componentes (CP1, CP2 e CP3) acumulam aproximadamente 79% da variabilidade total dos dados. Reter os três primeiros componentes para análises posteriores é bastante razoável para uma representação parcimoniosa das variáveis. A escolha de três componentes parece adequada tomando como referência a Tabela 3, verificando-se que, com três componentes principais, a menor porcentagem de variância individual explicada é a da variável VEV com 67%.

A interpretação das componentes será feita com base nas Tabelas 3, 4, 5 e 6 e Figs. 2, 3 e 4. A Tabela 4 apresenta os coeficientes das componentes principais. As variáveis NO, NOx, CO, O<sub>3</sub> e VEV possuem os maiores coeficientes na primeira componente principal, resultado já esperado, pois as variáveis foram padronizadas e as correlações entre as variáveis são altas e algumas próximas umas das outras. A segunda componente principal apresenta os maiores coeficientes para as variáveis NO<sub>2</sub>, RSG, TEM e UR.

Com auxílio da Tabela 3, é possível criar os mapas fatoriais apresentados nas Figs. 2, 3 e 4. A Tabela 5 apresenta a porcentagem explicada por cada variável e a Tabela 6 ilustra as contribuições de cada variável para a componente principal.

Vetores que representam pontos variáveis de alta contribuição, ou seja, extremos próximos à circunferência do círculo de correlações do mapa fatorial, representam as variáveis que justificam a maior dispersão. São essas as variáveis que desempenham um papel mais relevante na



**Figura 1** - Matrix de correlações entre as variáveis estudadas.

**Tabela 2** - Autovalores e explicação da variância total.

Componentes	Autovalor	Explicação (%)	Explicação acumulada
CP1	4,02	44,68	44,68
CP2	2,28	25,30	69,98
CP3	0,85	9,45	79,43
CP4	0,57	6,31	85,74
CP5	0,43	4,81	90,56
CP6	0,33	3,64	94,20
CP7	0,32	3,51	97,71
CP8	0,20	2,21	99,92
CP9	0,01	0,08	100,00

**Tabela 3** - Correlações entre as variáveis e as componentes principais.

Variável	CP1	CP2	CP3
NO <sub>2</sub>	0,36	-0,65	0,59
NO	0,83	-0,30	-0,41
NO <sub>x</sub>	0,84	-0,46	-0,19
CO	0,70	-0,40	-0,03
O <sub>3</sub>	-0,84	-0,08	0,08
VEV	-0,74	-0,10	-0,33
RSG	-0,50	-0,63	-0,36
TEM	-0,46	-0,76	0,20
UR	0,55	0,63	0,10

análise, pois são as variáveis determinantes da componente principal.

A análise do mapa fatorial apresentado na Fig. 2 evidencia que as variáveis NO, NO<sub>x</sub>, CO e NO<sub>2</sub> encontram-se positivamente correlacionadas, essa última de forma mais fraca em relação às outras três, pois sua representação no mapa não fica próxima ao círculo das

correlações. Esta correlação entre as variáveis NO, NO<sub>2</sub>, NO<sub>x</sub> e CO indica que estes poluentes possivelmente são oriundos das mesmas fontes, que no caso da localidade estudada são as fontes veiculares.

Também é possível observar que pelo inventário de emissões veiculares do INEA para a Região Metropolitana do Rio de Janeiro (2016) que utilizou o ano base de 2013, o

**Tabela 4** - Coeficientes das componentes principais.

Variável	CP1	CP2	CP3
NO <sub>2</sub>	0,18	-0,43	0,64
NO	0,41	-0,20	-0,45
NOx	0,42	-0,31	-0,21
CO	0,35	-0,27	-0,04
O <sub>3</sub>	-0,42	-0,05	0,09
VEV	-0,37	-0,07	-0,35
RSG	-0,25	-0,42	-0,39
TEM	-0,23	-0,50	0,22
UR	0,27	0,42	0,11

**Tabela 5** - Porcentagem explicada das variâncias individuais.

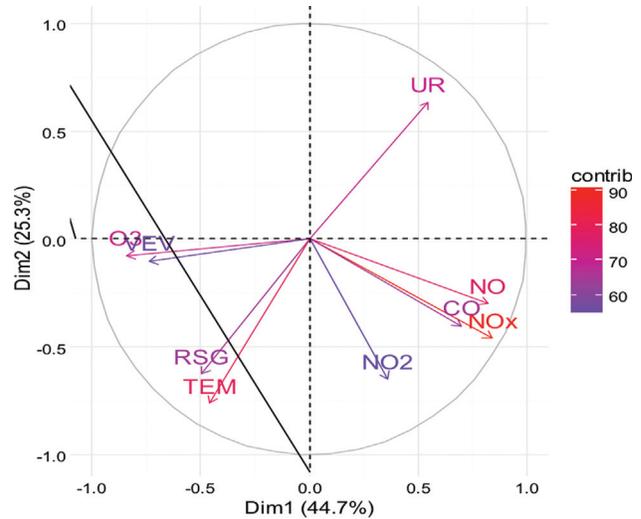
Variável	CP1	CP2	CP3	Soma
NO <sub>2</sub>	0,13	0,42	0,35	0,90
NO	0,68	0,09	0,17	0,94
NOx	0,71	0,21	0,04	0,96
CO	0,49	0,16	0,00	0,65
O <sub>3</sub>	0,71	0,01	0,01	0,73
VEV	0,55	0,01	0,11	0,67
RSG	0,25	0,39	0,13	0,77
TEM	0,21	0,58	0,04	0,83
UR	0,30	0,40	0,01	0,71

**Tabela 6** - Contribuição das variáveis nas componentes principais.

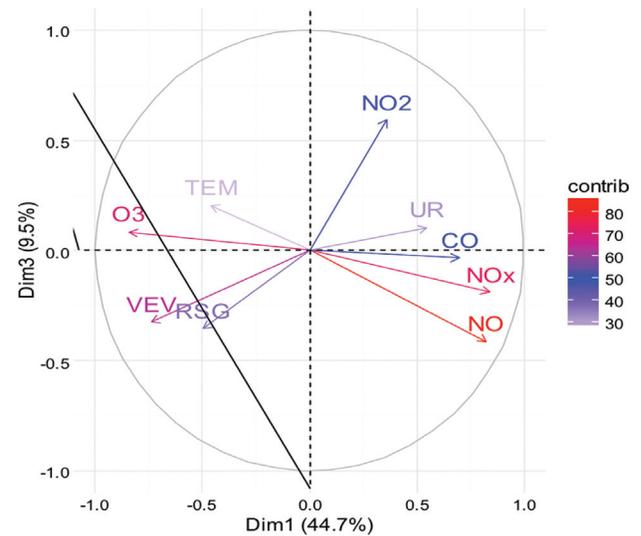
Variável	CP1	CP2	CP3
NO <sub>2</sub>	3,21	18,53	41,35
NO	16,95	3,98	19,94
NOx	17,57	9,31	4,22
CO	12,13	7,13	0,14
O <sub>3</sub>	17,64	0,25	0,80
VEV	13,60	0,47	12,50
RSG	6,12	17,26	14,86
TEM	5,31	25,40	4,93
UR	7,47	17,68	1,27

CO contribui com 55,61% das emissões, seguido de 32,73% dos NOx e frações menores dos demais poluentes, evidenciando a importância do CO e NOx.

Ozônio e VEV também são positivamente correlacionadas, mas as desse grupo apresentam correlações negativas com as do grupo anterior; RSG e TEM são positivamente correlacionadas e não apresentam correlação significativa com as do grupo (NO, NOx, CO e NO<sub>2</sub>), devido às quase ortogonalidades entre os vetores. Pode-se explicar que a alta correlação entre RSG e TEM, pois a radiação solar aquece a atmosfera e o solo em diferentes



**Figura 2** - Mapa Fatorial da Análise dos Componentes Principais (CP1 x CP2).



**Figura 3** - Mapa Fatorial da Análise dos Componentes Principais (CP1 x CP3).

comprimentos de onda e o solo aquecido retorna parte da energia na forma de radiação infravermelha também aquecendo a atmosfera. Uma correlação positiva entre VEV e O<sub>3</sub> pode indicar que este poluente pode estar sendo transportado de localidades vizinhas, sendo uma parte produzido localmente e outra parte produzido em regiões de maior produção de NOx e COV, seus precursores. Também é possível observar uma forte correlação entre O<sub>3</sub> e TEM, visto que o ozônio é formado por processos fotoquímicos dependentes da luz do sol e da temperatura. Outra observação oriunda da Fig. 2 é a correlação inversa entre TEM e UR. No início da manhã costuma-se observar baixos valores de TEM e altos valores de UR. Com o aumento da TEM ao longo do dia a água passa para a fase gasosa reduzindo o valor da UR e incrementando o valor da pressão atmo-

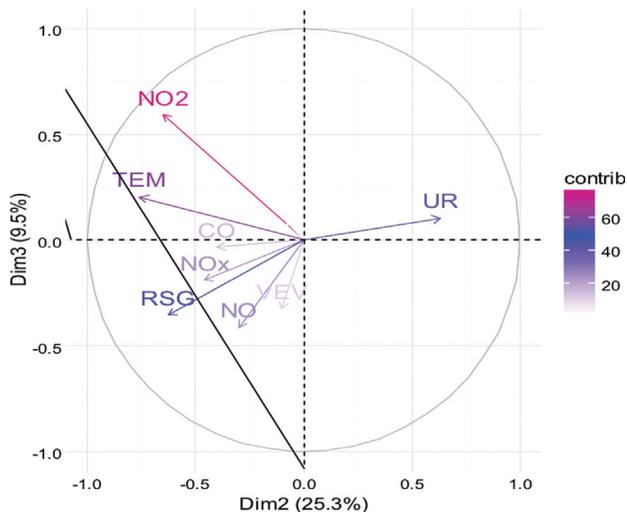


Figura 4 - Mapa Fatorial da Análise dos Componentes Principais (CP2 x CP3).

férica. No final do dia comportamento similar é observado, com a redução da TEM e incremento da UR. Análise semelhante pode ser feita nos mapas das Figs. 3 e 4.

Em geral a umidade relativa não apresenta correlação com as variáveis estudadas, a não ser com a temperatura. As moléculas estudadas neste trabalho são da fase gasosa e a UR não participa das reações de formação e decomposição. Caso o estudo fosse focado em espécies presentes na fase heterogênea da atmosfera, tais como dióxido de enxofre, sulfatos e material particulado de tamanho fino a UR teria provavelmente mais correlação.

Com base nas componentes principais extraídas, adotando-se a medida de distância euclidiana e o critério de Ward, procurou-se alocar as variáveis em uma quantidade de agrupamentos (*clusters*) homogêneos internamente e heterogêneos entre si. Na Fig. 5 é apresentado um dendograma com as distancias euclidianas com os clusters pro-

postos, que ilustram um resumo dos grupos formados. A variável UR foi alocada em um cluster individual; as variáveis (NO, NOx, CO e NO<sub>2</sub>) constituem um segundo *cluster* e as demais variáveis (O<sub>3</sub>, VEV, RSG, TEM) formam o terceiro *cluster*.

A Fig. 5 indica uma forte evidência que os poluentes primários NOx e CO possuem uma única contribuição das fontes veiculares locais, já que na localidade estudada não existem fontes fixas. Também é possível concluir que o ozônio, um poluente secundário não emitido por nenhuma fonte tem sua origem nos poluentes primários e os efeitos das variáveis meteorológicas, como a RSG, TEM e uma parcela de transporte pelos ventos. Pode se estimar que o ozônio seja tanto transportado pelos ventos como também pelo transporte dos seus precursores COVs e NOx e ser produzido localmente, já que a estação estava localizada a menos de 500 m da Auto Estrada Lagoa Barra, em uma localidade com características singulares, perto da Floresta da Tijuca, do Túnel Dois Irmãos, da Lagoa Rodrigo de Freitas e do Oceano Atlântico.

#### 4. Conclusões

Este estudo permitiu correlacionar as variáveis mensuradas por uma estação automática de monitoramento da qualidade do ar na cidade do Rio de Janeiro e desta forma auxiliar no entendimento da inter-relação das variáveis, em uma sinergia entre a química da atmosfera e as ferramentas estatísticas.

Foi possível entender que as fontes veiculares são prioritárias na região estudada, com alta correlação entre os poluentes primários CO e NO, emitidos diretamente pelas fontes móveis. Apesar dos COV não terem sido monitorados neste estudo o ozônio também apresentou uma correlação com o NOx, um dos precursores do ozônio, assim como a correlação deste com a radiação solar e temperatura. O ozônio também apresentou uma correlação com a veloci-

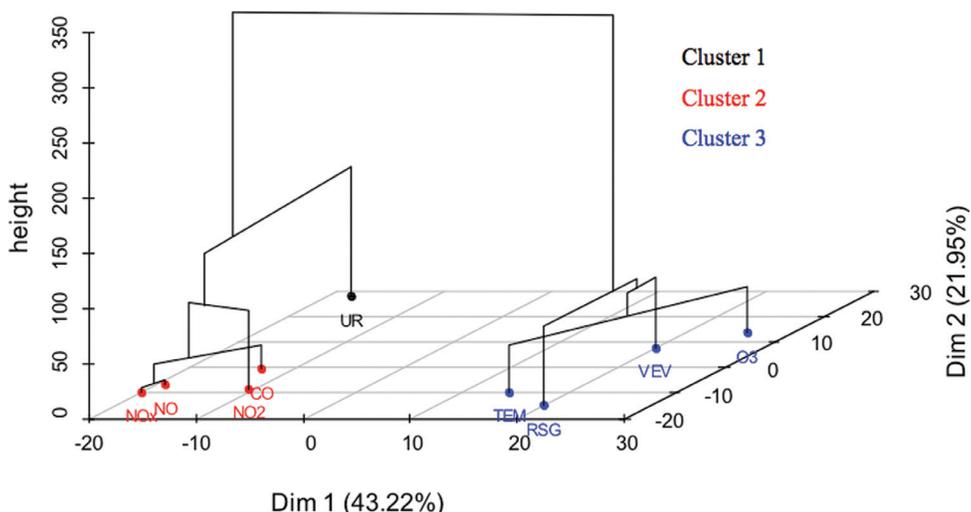


Figura 5 - Dendrograma das variáveis estudadas separadas por clusters.

dade do vento, o que pode indicar que está sendo transportado de outras localidades.

Estudos estão em desenvolvimento para outras localidades da cidade para um banco de dados mais robusto, que envolve diferentes anos e estações do ano, incluindo também os COV de uma forma global e alguns outros de forma especiada, para desta forma poder tentar chegar a um conjunto mínimo de dados a serem monitorados e conhecer a qualidade do ar.

Espera-se que este estudo e vindouros venham auxiliar as agências ambientais a tomarem decisões sobre a qualidade do ar e mesmo chegar a um conjunto mínimo de parâmetros a serem monitorados, reduzindo desta forma os elevados custos de se instalar e operar uma estação automática da qualidade do ar, que em geral são de 100 a 300 de milhares de dólares anuais.

### Agradecimentos

A SMAC pela cessão e uso dos dados da estação de monitoramento automática. A FAPERJ e ao CNPq pelo contínuo apoio às pesquisas do grupo.

### Referências

- ATKINSON, R. Atmospheric chemistry of VOCs and NOx. *Atmospheric Environment*, v. 34, p. 2063-2101, 2000.
- BARTLETT, M.S. Tests of Significance in Factor Analysis. *British Journal of Statistical Psychology*, v. 3, p. 77-85, 1950.
- FÁVERO, L.P.; BELFIORE, P. *Análise de Dados: Técnicas Multivariadas Exploratórias com SPSS e STATA*. Elsevier Editora, 2015.

- FINLAYSON-PITTS, B.J.; PITTS, J.N. *Chemistry of the Upper and Lower Atmosphere*. Academic Press, San Diego, 2000.
- HAIR J.R.; BLACK J.F.; BABIN, W.C.; ANDERSON, B.J.; TATHAM, R.E. *Análise multivariada de dados*. 6 ed., Bookman, Porto Alegre, 2009.
- INEA. *Inventário de Emissões de Fontes Veiculares: Região Metropolitana do Rio de Janeiro – Ano Base 2013*, 2016.
- LÊ, S.; JOSSE, J.; HUSSON, F. *FactoMineR: An R Package for Multivariate Analysis*. *Journal of Statistical Software*, v. 25, p. 1-18, 2008.
- LORA, E.E.S. *Prevenção e controle da poluição nos setores energéticos, industrial e de transportes*. Interciência, 2002.
- LUNA, A.S.; PAREDES, M.L.L.; OLIVEIRA, G.C.G.; CORRÊA, S.M. Prediction of ozone concentration in tropospheric levels using artificial neural networks and support vector machine at Rio de Janeiro, Brazil. *Atmospheric Environment*, v. 98, p. 98-104, 2014.
- MARTINS, E.M.; NUNES, A.C.L.; CORRÊA, S.M. Understanding Ozone Concentrations During Weekdays and Weekends in the Urban Area of the City of Rio de Janeiro. *Journal of the Brazilian Chemical Society*, v. 26, p. 1967-1975, 2015.
- NETO, J.M.M.; MOITA, G.C. *Uma introdução à análise exploratória de dados multivariados*. Scielo Editora, 1997.
- ORLANDO, J.P.; ALVIM, D.S.; YAMAZAKI, A.; CORRÊA, S.M.; GATTI, L.V. Ozone precursors for the São Paulo Metropolitan Area. *Science of the Total Environment*, v. 408, p. 1612-1620, 2010.
- R CORE TEAM 2013. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- TEIXEIRA, J.R.; SOUZA, C.V.; SODRÉ, E.D.; CORRÊA, S.M. Volatile Organic Compound Emissions from a Landfill, Plume Dispersion and the Tropospheric Ozone Modeling. *Journal of the Brazilian Chemical Society*, v. 23, p. 496-504, 2012.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License which permits unrestricted non-commercial use, distribution, and reproduction in any medium provided the original work is properly cited.