LVI SBPO SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL
DE 04 ATÉ 07 NOVEMBRO 2024
FORTALEZA | CEARÁ

ESTIMATING THE ASSOCIATION BETWEEN GREEN URBAN AREAS AND PM10 POLLUTION USING CONDITIONAL INFERENCE TREES

Leticia Melcop Sant'Anna
lmsantanna@hotmail.com

Henrique Schittine Campos
henriquecampos19@gmail.com

Leonardo dos Santos Lourenço Bastos
lslbastos@puc-rio.br

Soraida Aguilar
soraguilar@puc-rio.br

Department of Industrial Engineering (DEI)
Pontifical Catholic University of Rio de Janeiro – PUC-Rio
Rua Marquês de São Vicente, 225 | Bairro Gávea, Rio de Janeiro – RJ, 22451-900

ABSTRACT

High concentrations of PM10 cause negative health conditions and damage to city infrastructure, soil, and bodies of water, while trees play a crucial role in mitigating PM10; however, their true impact in urban environments remains unclear. This study aims to estimate the association between urban green areas and PM10 concentration levels by using Conditional Inference Tree model, to access which variables are most relevant to PM10 concentrations, including the green area and meteorological variables such as solar radiation, atmospheric pressure, humidity, wind speed and temperature. The dataset used is from air monitoring stations in city of Rio de Janeiro from 2011 to 2022, totaling 832 valid measurements. Our findings show that while green urban areas are associated with the concentration of PM10, the variables of station, month of the year, air pressure, wind speed and solar radiation have a higher importance to modifying the concentration.

KEYWORDS. Air Quality. Particulate Matter. Green Areas.

Paper topics AG&MA – OR in Agriculture and Environment; BDA – Big Data and Analytics

https://proceedings.science/p/193719?lang=pt-br

1. Introduction

Air pollution is recognized as the most prominent environmental hazard to human health due to its significant contribution to diseases [WHO 2021]. Particulate Matter smaller than 10 microns (PM10) is especially notable for its role in developing a plethora of respiratory and cardiovascular conditions after long-term exposure [WHO 2021]. Aside from the health risks associated with PM10, the accumulation of these particles can cause other maleficent effects, such as acidification of bodies of water, hindering the photosynthesis of plants, acceleration of the corrosion of metal and degradation of concrete [Luoma & Rainbow 2008].

One of the ways to diminish concentrations of PM10 is through trees and green spaces due to their capability of intercepting particles. They play a crucial role in improving air quality by removing pollutants from the atmosphere, acting as natural filters, and capturing and retaining particulate matter on their surfaces [EPA 2015]. This situation presents an interesting dichotomy: trees are both important sinkholes of PM10, but they are negatively affected by their concentration. It creates the question of whether they are a relevant way to mitigate PM10 concentrations in urban areas, or even though, despite the green area's presence in urban regions, they can play their role as natural filters, or contrary to what one might think, this compensatory effect is not existing due to high PM10 emissions.

Hence, this study aims to assess the association between green spaces and the presence of PM10. We used a machine-learning-based model (the conditional inference tree) to explore the associations among explainable variables (green areas, meteorological variables and temporal factor) with the response (PM10) in data from Rio de Janeiro, Brazil.

2. Methodology

We investigated the association between PM10 concentration levels and green urban areas in Rio de Janeiro, one of Brazil's largest urban centers that contains high amounts of PM10 producers. We also included as explainable variables temperature, relative humidity, solar radiation, atmospheric pressure and wind speed, and the green space as potential features related with PM10 [Zhang et al. 2023].

The dataset used for the analyses was retrieved from the Air Quality monitoring program from the City Hall of Rio de Janeiro (*MonitorAR Rio)* [Data.Rio 2023]. It measures the concentration of PM10 ($\mu g/m^3$) in the air on a hourly frequency, in eight locations of the city, as well as the wind speed (m/s), wind direction (°), solar radiation (w/m^2), rain precipitation (mm), relative humidity (%), temperature (°C), air pressure (mbar), and the location's latitude and longitude. 707,093 measurements across 11 years (2011-2022) were obtained. Data on the green area was obtained from a report of the City Hall of Rio de Janeiro [City Hall of Rio de Janeiro 2018] and consisted of the proportion of green areas per neighborhood in the city in 2014 and the variation between 2010 to 2016 in squared meters of vegetation in each city planning area (AP), from which we estimated the month proportion of green area in the whole period. A few inconsistencies and all the entries with outlier values were removed. Then, we obtained the data at a month-year and station resolution, by estimating the monthly average of the measurements per location. The final data set used for the analysis with 832 monthly records.

To model the association of the explainable variables and the response variable PM10, we applied the Conditional Inference Trees (CIT) method. CIT is a machine learning model based on decision trees; however, each tree is created employing recursive partitioning process considering conditional inference procedures instead of prediction, which allows for unbiased inference of associations [Hothorn et al. 2015]. To evaluate the association of the variables with PM10, especially the green area, we performed a one-step analysis. The data cleaning process and data preparation was supported by using the programming language *R*. The programming code of data analysis is available in GitHub repository [https://github.com/MelcopLeticia/PM10-and-Green-Area].

3. Results & Discussion

When describing the distribution of each variable individually by station, we observed that the PM10 concentration values are similar among the stations, except for the AV (Copacabana), where it is higher. Additionally, the values for humidity, temperature and pressure differ little between stations. The percentage of green areas covering each neighborhood are different from each other, with the median being the value defined for December 2014. Lastly, the solar radiation variable differs between the stations, which could be due to the surrounding buildings preventing sunlight from reaching the monitoring stations.

We generated 10 CIT trees, for all meteorological variables considering one tree for each station and two trees for all stations with and without the variable month. The variables that caused splits in each tree are displayed in the marked cells on **Table 1**.

Table 1 – Analysis of variables associated with PM10 in CIT generated for each station

Station	Green Area	Month	Humidity	Air Pressure	Temp.	Wind Speed	Solar Rad.
AV						X	
BG				X			
CA	X				X		
CG			X	X			
IR			X	X		X	
PG				X			
SC	X	X			X	X	
SP	X	X					X
All Stations*		X	X			X	X
All Stations**	X	------	X	X		X	X

All Stations* refers to the CIT model with Month and Station as relevant splitting variables and All Stations** refers to the tree with only Meteorological variable and Green Area.

The Machine Learning algorithm shows that green areas seem to have a relation with PM10 levels. However, when spatial and temporal variables are added, they appear to have a better association with PM10. In the splits per station, green areas also appeared as a significant variable.

An important discussion is about the spatial and temporal relationship of the data with PM10 concentration. The variable green area only appears in 40% splits of the trees generated. A possible explanation for this can be the presence of seasonal data and specific factors about the station's location that were not fully captured in our data sets. Another essential aspect is the split with all variables, in which the month appeared as a relevant one. It indicates that seasonality can significantly explain PM10 concentration in a way that the meteorological variables modeled could not by themselves.

The results also indicate that these meteorological variables may have a more significant impact on PM10 levels than green areas. Wind speed, humidity, solar radiation, and atmospheric pressure are more associated with PM10 concentrations. According to the CETESB [2024], meteorological conditions are very correlated with pollutants concentration in the air. Higher wind speeds enhance the pollutants dispersion, increased humidity can remove PM10 particles in the air, changes in pressure can impact air circulation creating a thermic inversion, a phenomenon that imprisons the pollutants in the air [CETESB 2024]. There are a few other limitations in our project that can affect this result. The green area data used in the study were extrapolated based on a temporal average of vegetation cover variation between 2010 and 2016. Moreover, the extrapolation assumes that the monthly variation was the same, which may not reflect the reality of urban green areas and not even the real amount of vegetation in each region nowadays.

5. Conclusion

The study provides insights into the relation between PM10 levels in Rio de Janeiro, highlighting the complex interaction of various factors influencing air quality. The location and seasons have a much more correlation to PM10 than green areas. The data also indicates that no PM10 level in Rio is critical, according to CETESB standards. This article also suggests the

necessity of a more holistic approach that incorporates various environmental and socio-economic factors in the management of air quality.

References

City Hall of Rio de Janeiro (2018.) Programa de Monitoramento da Cobertura Vegetal. Subsecretaria de Meio Ambiente. Web page: https://www.rio.rj.gov.br/dlstatic/10112/6438610/4221813/75ProgramadeMonitoramentodaCoberturaVegetal.pdf. Acessado: 2024-04-18.

Companhia de Tecnologia de Saneamento Ambiental (CETESB) (2024). Poluentes, São Paulo. Web page: https://cetesb.sp.gov.br/ar/poluentes/. Acessado: 2024-04-24.

Data.Rio (2023). Dados horários do monitoramento da qualidade do ar – MonitorAr. Web page: https://www.data.rio/datasets/PCRJ::dados-hor%C3%A1rios-do-monitoramento-da-qualidade-do-ar-monitorar/about. Acessado: 2024-03-24.

EPA (2015). Percent Particulate Matter (PM10) Removed Annually by Tree Cover. EnviroAtlas. Web page: https://www.epa.gov/enviroatlas. Acessado: 2024-05-05.

Hothorn, T., Hornik, K. e Zeileis, A (2015). ctree: Conditional inference trees. *The comprehensive R archive network*, *8*, 1-34.

Luoma, S. N. e Rainbow, P. S. (2008). Metal Contamination in Aquatic Environments: Science and Lateral Management. Cambridge University Press.

WHO (2021). Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. Geneva: World Health Organization.

Zhang, M., Chen, S., Zhang, X., Guo, S., Wang, Y., Zhao, F., Chen, J., Qi, P., Lu, F., Chen, M. e Bilal, M. (2023). Characters of particulate matter and their relationship with meteorological factors during winter Nanyang 2021–2022. *Atmosphere*, 14(1), 137.