Prediction of air pollution events in São Paulo based on surface meteorological variables

Andre Gomes Bessa Miranda[1] and Luciana Varanda Rizzo[2]

[1]Universidade Federal de São Paulo - UNIFESP
[2]Universidade de São Paulo - USP

November 21, 2022

Abstract

Large urban centers like the Metropolitan Region of São Paulo (MASP) are impacted by air pollution, especially by Inhalable particle matter (PM10). Persistent exceedance events (PEE) are defined as exceedance events that last for many consecutive days and occur simultaneously at many air quality monitoring stations across the MASP. This study aims to develop a predictive model for the occurrence of PEE in the MASP based on surface meteorological variables. Hourly PM10 concentrations from 12 air quality monitoring stations in the MASP between 2005 and 2021 were provided by the São Paulo State Environmental Agency (CETESB). Daily data on surface meteorological variables were provided by the IAG/USP meteorological station. Persistent exceedance events (PEE) were identified using the criteria: exceedance events that occurred simultaneously in at least 50% monitoring stations, persisting for at least 5 consecutive days. PEE occurrence was represented as a timeseries of a binary variable. The resulting daily dataset had 6204 lines and 13 attributes, without missing values. The dataset was divided into a training set (80%) and a test set (20%). A logistic regression model was applied, having the PEE occurrence (positive = 1) as the target value. The Variance Inflation Factor and the Stepwise Feature Selection method was applied to obtain an optimized subset of predictors. Model accuracy was accessed by the ROC curve and by a confusion matrix. Results indicate that PEE can be satisfactorily predicted by surface meteorological variables using a logistic regression. As for the next steps, we intend to extract easy-tocommunicate classification rules, aiming to support the development of warnings systems for air quality poor conditions in the MASP.

Prediction of air pollution events in São Paulo based on surface meteorological variables

MIRANDA, A. G. B. [1] ; RIZZO, L. V. [2]

[1] Undergrad in Environmental Sciences - Universidade Federal de São Paulo

[2] Laboratório de Física Atmosférica - Universidade de São Paulo

Email: [1] andre.miranda@unifesp.br, [2] lrizzo@usp.br

INTRODUCTION

Large urban centers like the Metropolitan Region of São Paulo (MASP) are impacted by air pollution, especially by Inhalable particle matter (PM_{10}) [1]. Air quality relies on several factors, like meteorological conditions and strength of emission sources. An exceedance event occurs when air pollutant concentrations exceed the air quality standards. Persistent exceedance events (PEE) are defined as exceedance events that last for many consecutive days and occur simultaneously at many air quality monitoring stations across the MASP [2]. This study aims to develop a predictive model for the occurrence of PEE in the MASP based on surface meteorological variables.

METHODS

Hourly PM_{10} concentrations from 12 air quality monitoring stations in the MASP (Fig. 1) between 2005 and 2021 were provided by the São Paulo State Environmental Agency (CETESB). Daily data on surface meteorological variables were provided by the IAG/USP meteorological station. Persistent exceedance events (PEE) were identified using the criteria [2]: exceedance events that occurred simultaneously in at least 50% monitoring stations, persisting for at least 5 consecutive days. PEE occurrence was represented as a timeseries of a binary variable. The resulting daily dataset had 6204 lines and 13 attributes, without missing values. The dataset was divided into a training set (80%) and a test set (20%). A logistic regression model was applied, having the PEE occurrence (positive = 1) as the target value. The Variance Inflation Factor and the Stepwise Feature Selection method was applied to obtain an optimized subset of predictors. Model accuracy was accessed by the ROC curve and by a confusion matrix.

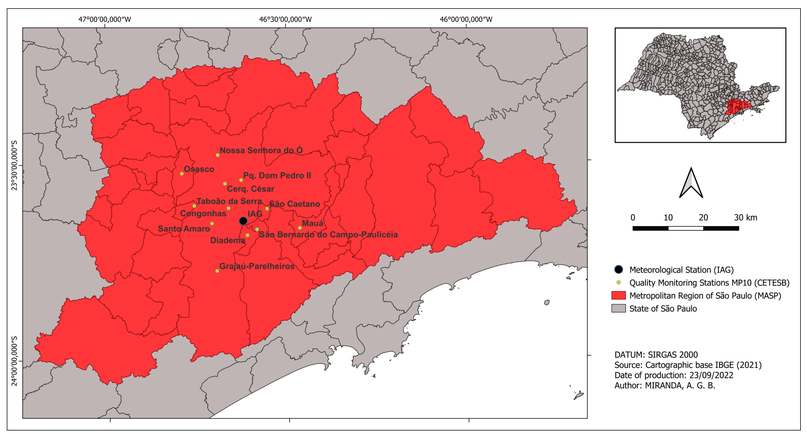


Fig. 1: Location of 12 air quality monitoring stations in the MASP.

RESULTS

Between 2005 and 2021, 122 PEE were identified in the MASP, lasting for 5 to 25 days. The events were more frequent in the fall and winter (Fig. 2). The events had a prevalence of 15.3%.

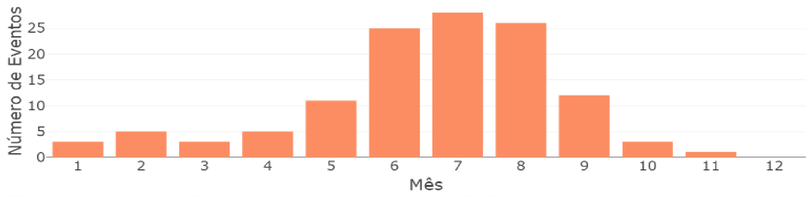


Fig.2: Occurrence of Persistent Exceedance Events (PEE) along the year.

During the events, 24h PM_{10} median concentrations increased by 67% compared to the preceding days, reaching 65 $\mu g.m^{-3}$, above the WHO air quality standard. There were substantial changes in the surface meteorological variables during the events, with an increase in Tmax and a decrease in RH (Fig. 3). Those changes are compatible with the occurrence of unfavorable conditions for air pollution dispersion.

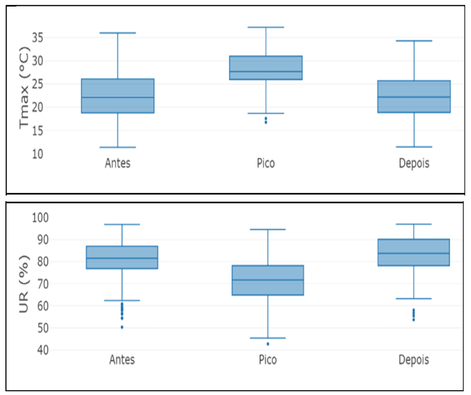
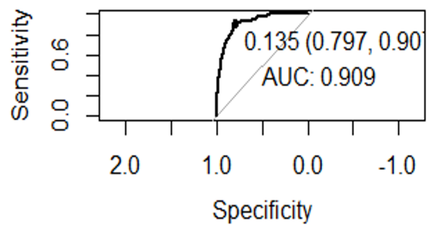


Fig. 3: Boxplot diagrams for meteorological variables before, during and after the PEE: daily maximum temperature (Tmax) and relative humidity (RH)

A logistic regression model was developed using 7 out of 13 available meteorological variables as predictors and a categorical variable for the season. All coefficients were significative with p<0.01. The area under the ROC curve was 91%, with the best threshold in 0.135 (Fig. 4a). Cross-validation for the test set showed a model accuracy of 81%, with sensibility of 0.91 and specificity of 0.80 (Fig. 4b).

a)



b)

		Reference	
		PEE =0	PEE =1
Prediction	PEE=0	819	19
	PEE =1	208	186

Fig. 4: ROC curve (a) and confusion matrix (b) for the test set.

The model coefficients and corresponding odds ratios (Table 1) indicate that Tmax and RH were the most relevant meteorological variables concerning PEE occurrence. An increase of 1 standard deviation in Tmax (4.6 ºC) was associated with a PEE chance increase of 1.98. On the other hand, a decrease of 1 standard deviation in RH (8.6%) resulted in a PEE chance increase of 72%. The seasons also showed a strong association wtih PEE occurrence, so that the chance is 10 times higher in the winter compared to summer.

	Units	σ	OR	2.5%	97.5%
Intercept	--	--	0.02	0.02	0.03
FAL	--	--	2.42	1.58	3.77
SPR	--	--	3.41	2.24	5.27
WIN	--	--	10.21	6.37	16.64
ustar	m/s	0.11	0.82	0.71	0.93
wind	m/s	1.8	0.53	0.47	0.61
irrad	W/m^2	6.3	0.43	0.35	0.53
precip	mm/day	11	0.81	0.63	1.01
press	mbar	3.5	1.66	1.43	1.92
Tmax	ºC	4.6	1.98	1.65	2.39
RH	%	8.6	0.28	0.24	0.33

Table 1: standard deviation (σ), odds ratio (OR) and confidence intervals for the predictive variables: seasons (SUM, FAL, WIN, SPR), friction (ustar) and wind velocity, irradiation, precipitation, pressure, maximum temperature (Tmax) and relative humidity (RH).

CONCLUSION

Results indicate that PEE can be satisfactorily predicted by surface meteorological variables using a logistic regression. As for the next steps, we intend to extract easy-to-communicate classification rules, aiming to support the development of warnings systems for air quality poor conditions in the MASP.

References:

[1] Andrade et al., 2017, Atmos. Env.,159, 66–82. https://doi.org/10.1016/j.atmosenv.2017.03.051.

[2] Oliveira et al., 2021, Int. J. Env. Sci. Tech. https://doi.org/10.1007/s13762-021-03778-1

Acknowledgements:

We thank CETESB and Estação Meteorológica do IAG for providing the data.

We thank FAPESP (2021/14342-7) for the undergrad research grant.