Statistical Methods to Forecast Air Quality in Taipa Ambient and Taipa Residential of Macao

Man Tat Lei, Joana Monjardino, Luisa Mendes, David Gonçalves, and Francisco Ferreira

Abstract Air pollution is a major concern issue on Macao since the concentration levels of several of the most common pollutants are frequently above the internationally recommended values. The low air quality episodes impacts on human health paired with highly populated urban areas are important motivations to develop forecast methodologies in order to anticipate pollution episodes, allowing establishing warnings to the local community to take precautionary measures and avoid outdoor activities during this period. Using statistical methods (multiple linear regression (MLR) and classification and regression tree (CART) analysis) we were able to develop forecasting models for the main pollutants (NO_2, $PM_{2.5}$, and O_3) enabling us to know the next day concentrations with a good skill, translated by high coefficients of determination (0.82–0.90) on a 95% confidence level. The model development was based on six years of historical data, 2013 to 2018, consisting of surface and upper-air meteorological observations and surface air quality observations. The year of 2019 was used for model validation. From an initially large group of meteorological and air quality variables only a few were identified as significant dependent variables in the model. The selected meteorological variables included geopotential

M. T. Lei (✉) · L. Mendes
Department of Sciences and Environmental Engineering, NOVA School of Science and Technology, NOVA University Lisbon, Lisbon, Portugal
e-mail: lei.man.tat@usj.edu.mo; l.tat@campus.fct.unl.pt

L. Mendes
e-mail: lc.mendes@fct.unl.pt

J. Monjardino · F. Ferreira
Center for Environmental and Sustainability Research, NOVA School of Science and Technology, NOVA University Lisbon, Lisbon, Portugal
e-mail: jvm@fct.unl.pt

F. Ferreira
e-mail: ff@fct.unl.pt

M. T. Lei · D. Gonçalves
Institute of Science and Environment, University of Saint Joseph, Macau, China
e-mail: david.goncalves@usj.edu.mo

C. Mensink and V. Matthias (eds.), *Air Pollution Modeling and its Application XXVII*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-662-63760-9_25

height, relative humidity and air temperature at different altitude levels and atmospheric stability characterization parameters. The air quality predictors used included recent past hourly levels of mean concentrations for NO_2 and $PM_{2.5}$ and maximum concentrations for O_3. The application of the obtained models provides the expected daily mean concentrations for NO_2 and $PM_{2.5}$ and maximum hourly concentrations O_3 for the next day in Taipa Ambient air quality monitoring stations. The described methodology is now operational, in Macao, since 2020.

1 Introduction

The World Health Organization has last update Air Quality Guidelines (AQG) in 2005, which sets the recommended threshold levels for each pollutant. Also, Macao has an air quality problem, in particular of high levels of nitrogen dioxide (NO_2), particulate matter ($PM_{2.5}$), and ozone (O_3), which often overstep the guidance values of AQG. In addition, there are a lot of studies that show that exposure to NO_2, $PM_{2.5}$, and O_3 have increased hospital admissions and emergency room visits and even led to death from heart or lung diseases in extreme cases (WHO, 2003). Thus, it is extremely important to develop a reliable air quality forecast for the concentration of NO_2, $PM_{2.5}$, and O_3 in Macao, which can alert the local population to take precautionary measures in case of a pollution episode (Neto et al., 2009). Figures 1, 2 and 3 showed the comparison of different air quality standards amongst WHO, EU, US, China, Macao, and Hong Kong for NO_2, $PM_{2.5}$, and O_3 (MEE, 2012; SMG, 2019; WHO Europe, 2006).

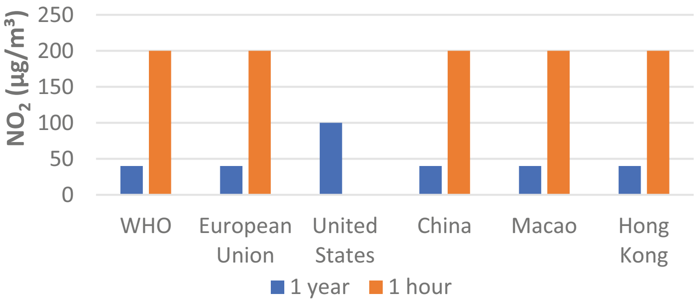

Fig. 1 Comparison of air quality standard for NO_2

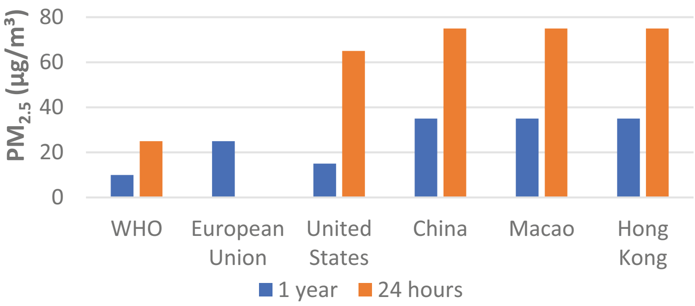

Fig. 2 Comparison of air quality standard for $PM_{2.5}$

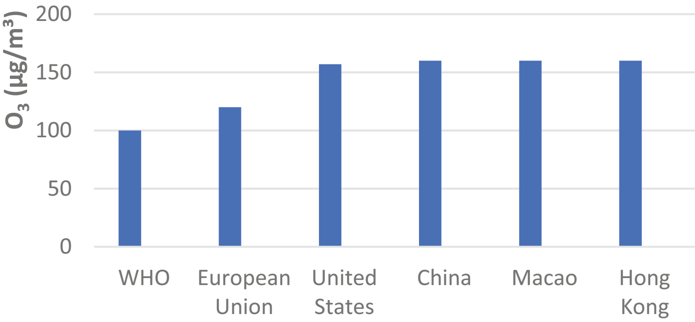

Fig. 3 Comparison of air quality standard for O_3

2 Methodology

To now the air quality for the next day was used statistical methods that were based on past data series analysis. For this paper was utilized multiple linear regression (MLR) and classification and regression tree (CART) analysis. As showed in previous work (Cassmassi, 1997), statistical models based on MLR and CART analysis were developed to forecast the average daily concentration for NO_2 and $PM_{2.5}$, and the maximum hourly O_3 levels for the next day, for the air quality monitoring stations of Taipa Ambient and Taipa Residential. Taipa Ambient is an ambient station, also a background representative station, which set the baseline for the levels of pollutant concentration. This station is located at Taipa Grande, the headquarter of Macao Meteorological and Geophysical Bureau (SMG). Taipa Residential is a high-density residential area station located in Taipa. This station is located at the Taipa Central Park, a leisure area for the local residents. A six-year period from 2013 to 2018 was selected as the period to develop the models, while the year of 2019 was selected for validation of the model.

Figure 4 shows the flowchart of the model development for the air quality forecast using statistical methods. The development of statistical model consists of collecting

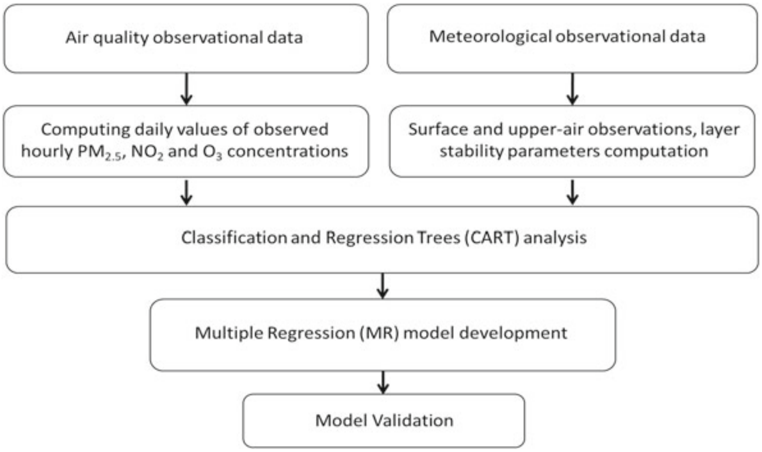

Fig. 4 Flowchart for the development of statistical air quality forecast models

the air quality data and meteorological data, followed by computing the hourly air quality data into the daily concentrations of NO_2, $PM_{2.5}$, and O_3. In addition, the meteorological data required to develop the statistical model would be extracted from different meteorological observations. These processed data would be analyzed by statistical methods such as multiple linear regression (MLR) and classification and regression trees (CART) analysis. The final procedure is to perform a model validation to ensure the accuracy of the next-day air quality forecast.

3 Results and Discussion

Table 1 provides the final list of meteorological and air quality parameters used as predictors, for each pollutant, in the obtained multiple regression models for the air quality air quality monitoring stations of Taipa Ambient and Taipa Residential. Table 1 shows $PM_{2.5}$ has a highly correlated relationship with past concentration

Table 1 Variables used in statistical model

Station	Pollutants	Variables used in equations
Taipa ambient	$PM_{2.5}$	PM_{25}_16D1, H_850, HRMD
	NO_2	NO_2_16D1, H_850, STB_925
	$O_{3\ MAX}$	$O_{3\ MAX}$_16D1, $O_{3\ MAX}$_23D1, H_850. HRMN
Taipa residential	$PM_{2.5}$	PM_{25}_16D1, TD_MD, TAR_925
	NO_2	NO_2_16D1, H_850, TD_MD
	$O_{3\ MAX}$	$O_{3\ MAX}$_16D1, $O_{3\ MAX}$_23D1, H_850, HRMN

levels (PM_{25}_16D1 as the average of the hourly values(μg/m^3) between 16:00 of yesterday and 15:00 of today) for both Taipa Ambient and Taipa Residential air quality monitoring stations, geopotential height (m) at 850 hPa (H_850), and average relative air humidity (%) (HRMD) for Taipa Ambient, average dew point temperature (°C) (TD_MD), and air temperature (°C) at 925 hPa (TAR_925) for Taipa Residential. In addition, NO_2 has a highly correlated relationship with past concentration levels NO_2_16D1and geopotential height (m) at 850 hPa for both Taipa Ambient and Taipa Residential air quality monitoring stations, atmospheric stability (°C) at 925 hPa (STB_925) for Taipa Ambient, and average dew point temperature (°C) (TD_MD) for Taipa Residential. Furthermore, O_3 has a highly correlated relationship with past concentration levels $O_{3\,MAX}$_16D1, $O_{3\,MAX}$_23D1 (as the maximum hourly values (μg/m^3) between 00:00 and 23:00 of yesterday), geopotential height (m) at 850 hPa and minimum relative air humidity (%) (HRMN) for both Taipa Ambient and Taipa Residential air quality monitoring stations.

An example of one of the regression equations delivered is the following for next-day 24 h-average NO_2 at Taipa Ambient:

$$NO_2 = (0.914 \times NO_2\_16D1) + (0.004 \times H\_850) - (0.734 \times STB\_925) \quad (1)$$

Figure 5 shows the graph of the model validation for the MLR models of NO_2 concentrations in Taipa Ambient air quality monitoring station in 2019.

Table 2 shows the model performance indicators of $PM_{2.5}$, NO_2, and $O_{3\,MAX}$ for Taipa Ambient and Taipa Residential air quality monitoring stations. The model performance indicators include coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and BIAS. The results obtained from MLR and CART models perform a coefficient of determination between 0.86 and 0.87 for Taipa Ambient and between 0.78 and 0.88 for Taipa Residential. The air

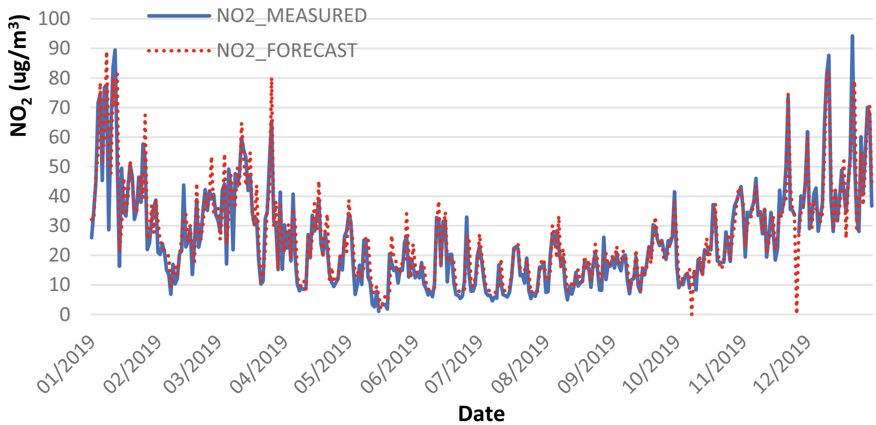

Fig. 5 Observed and predicted NO_2 concentrations value using MLR models in Taipa Ambient (2019)

Table 2 Model performance indicators

Station	Pollutant	Performance indicators				Model built using CART or MR	
		R^2	RMSE (μg/m^3)	MAE (μg/m^3)	BIAS (μg/m^3)	CART	MR
Taipa ambient	$PM_{2.5}$	0.86	4.8	3.1	0.2		✓
	NO_2	0.87	6.1	4.2	1.0		✓
	$O_{3\,MAX}$	0.86	23.7	14.7	−1.6	✓	✓
Taipa residential	$PM_{2.5}$	0.88	5.6	3.5	−0.1		✓
	NO_2	0.87	5.6	4.1	0.6		✓
	$O_{3\,MAX}$	0.78	20.9	12.7	1.3	✓	✓

quality forecast is at best forecasting the levels of NO_2 concentration in Taipa Ambient and is at best forecasting the levels of $PM_{2.5}$ concentration in Taipa Residential. All of the statistical models were built using MLR, while the models for the maximum hourly ozone were built using both MLR and CART.

The ambient station was better at predicting NO_2 while the residential station was better at predicting the maximum hourly concentration of O_3. Also, the developed models provide a better understanding of different air quality and meteorological variables and also the relationship between these variables. Furthermore, the variable that explained most of the variability is the 16D1 concentration for NO_2, $PM_{2.5}$, and O_3.

4 Conclusion

The work presented is an air quality forecast using statistical methods, based on a detailed analysis of both air quality and meteorological variables for NO_2, $PM_{2.5}$, and O_3. The final objective of this study is to develop a daily air quality forecast using statistical methods to predict the daily average of NO_2, $PM_{2.5}$, and maximum hourly O_3 levels for the next day, in the Taipa Ambient air quality monitoring station (background location) and the Taipa Residential air quality monitoring station (the high-density residential location). The models for NO_2, $PM_{2.5}$ and O_{3_MAX} used independent variables including the average of the hourly values between 16:00 of yesterday and 15:00 of today for NO_2, $PM_{2.5}$ and $O_{3\,MAX}$ respectively, the average of the hourly values between 00:00 and 23:00 of yesterday for $O_{3\,MAX}$, geopotential height at 850 hPa, atmospheric stability at 850 hPa and 925 hPa, air temperature at 925 hPa, average dew point temperature, minimum relative air humidity and average relative air humidity. The use of statistical models was successful in forecasting the average daily concentrations with MLR for NO_2 and $PM_{2.5}$ and MLR and CART analysis for the peak levels for maximum hourly O_3 for next day and be able to forecast the high concentration of pollution episodes, for both Taipa Ambient and Taipa Residential.

Acknowledgements The work developed was supported by The Macao Meteorological and Geophysical Bureau (SMG). The research work of CENSE is financed by Fundação para a Ciência e Tecnologia, I.P., Portugal (UID/AMB/04085/2019).

References

Cassmassi, J. C. (1997). Objective ozone forecasting in south coast Air Basin: Updating the objective prediction models for the late 1990's and Southern California ozone study (SCOS97-NARSTO) applications.

MEE. (2012). Ambient air quality standards. Retrieved from http://210.72.1.216:8080/gzaqi/Document/gjzlbz.pdf.

Neto, J., Torres, P., Ferreira, F., & Boavida, F. (2009). Lisbon air quality forecast using statistical methods. *International Journal of Environment and Pollution, 39*(3/4), 333. https://doi.org/10.1504/IJEP.2009.028695

SMG. (2019). Resumo anual sobre qualidade do ar em Macau—2019. Retrieved from https://cms.smg.gov.mo/uploads/sync/pdf/AIR_report/p_IQA_annual_report/IQA_2019.pdf.

WHO. (2003). Health aspects of air pollution with particulate matter, ozone and nitrogen dioxide. *Health Aspects of Air Pollution with Particulate Matter, Ozone and Nitrogen Dioxide*, (January), 98. https://doi.org/10.2105/AJPH.48.7.913.

WHO Europe. (2006). Air quality guidelines. Retrieved from http://202.171.253.71/www.euro.who.int/__data/assets/pdf_file/0005/78638/E90038.pdf.