

## MACAO AIR QUALITY FORECAST USING STATISTICAL METHODS

MAN TAT LEI<sup>1,3</sup>, JOANA MONJARDINO<sup>2</sup>, LUISA MENDES<sup>1</sup> & FRANCISCO FERREIRA<sup>2</sup>

<sup>1</sup>Department of Sciences and Environmental Engineering, NOVA School of Science and Technology,  
NOVA University Lisbon, Portugal.

<sup>2</sup>Center for Environmental and Sustainability Research, NOVA School of Science and Technology,  
NOVA University Lisbon, Portugal.

<sup>3</sup>Institute of Science and Environment, University of Saint Joseph, Macau, China.

### ABSTRACT

The levels of air pollution in the cities of Greater Bay Area in Southern China, including Macao, are extremely high and often exceeded the levels recommended by World Health Organization Air Quality Guidelines. In order for the population to take precautionary measures and avoid further health risks under high pollutant exposure, it is important to develop a reliable air quality forecast. Statistical models based on multiple regression analysis were developed successfully for Macao to predict the next-day concentrations of particulate matter ( $PM_{10}$  and  $PM_{2.5}$ ) for Taipa Ambient, a background representative station located within the area of Macao ( $32.9 \text{ km}^2$ ), at Taipa Grande, the headquarter of Macao Meteorological and Geophysical Bureau. The two developed models were statistically significantly valid, with a 95% confidence level with high coefficients of determination. A wide range of meteorological and air quality variables were identified, and only some were selected as significant dependent variables. The meteorological variables such as geopotential height and relative humidity at different vertical levels were selected from an extensive list of variables. The air quality variables that translate the resilience of the recent past concentrations of each pollutant were the ones selected. The models were based in meteorological and air quality variables with five years of historical data, from 2013 to 2017. The data from 2013 to 2016 were used to develop the statistical models and data from 2017 were used for validation purposes, with high coefficients of determination between predicted and observed daily average concentrations (0.92 and 0.89 for  $PM_{10}$  and  $PM_{2.5}$ , respectively). The results are expected to be the basis for an operational air quality forecast for the region.

*Keywords:* air pollutants, air quality forecast, management, modelling, monitoring.

### 1 INTRODUCTION

Macao is located in Southern China and one of the cities within the Greater Bay Area. The historical center of Macao has been awarded ‘World Cultural Heritage’ status by the United Nations Educational, Scientific and Cultural Organizations (UNESCO) in 2005. Nevertheless, the land area of Macao is extremely limited due to rapid growth of population and lack of land resources, and thus, the United Nations World Prospects Report had listed Macao as the number one most densely populated region in the world [1]. Macao has a land area of 30.8 square kilometers with a population of 653,100, which leads to a population density of 21,100 per square kilometers [2]. Due to the high population density, the health impact of air pollution is significant in Macao. The levels of particulate matter (PM) in Macao and its neighboring cities in the Greater Bay Area are extremely high and often exceeding the established limit values recommended by World Health Organization (WHO) Air Quality Guidelines (AQG). Therefore, it is important to develop a reliable prediction methodology for the concentration of PM, which can provide alert for health hazards in advance. Small particles less than 10 micrometers in diameter pose the greatest problems, specifically due to the fine particles below 2.5 micrometers that can get deep into the respiratory system, and some may even get into the bloodstream. Exposure to such PM can affect cardiovascular system. People with heart or lung diseases, older adults and children are considered at greater

Table 1: Air quality standards set at different countries and/or by different institutions [5]–[11].

Source	PM <sub>10</sub> ( $\mu\text{g}/\text{m}^3$ )		PM <sub>2.5</sub> ( $\mu\text{g}/\text{m}^3$ )	
	1 year	24 hours	1 year	24 hours
WHO	20	50	10	25.0
European Union	40	50	25	
United States	50	150	15	65.0
China (Classes 1)	40	50	15	35.0
China (Classes 2)	70	150	35	75.0
Macao	70	150	35	75.0
WHO Interim Target-1	70	150	35	75.0
WHO Interim Target-2	50	100	25	50.0
WHO Interim Target-3	30	75	15	37.5

risk from the exposure to PM. Numerous studies showed that exposure to PM has increased hospital admissions and emergency room visits and leading even to death from heart or lung diseases [3]. The exposure to PM would increase the chance of hospital admissions for cardiovascular and respiratory disease and mortality in the world [4].

The Macao Meteorological and Geophysical Bureau (SMG) adopted the WHO interim target-1 (IT-1) for the threshold of pollutants, which has a less strict standard on the pollutants compared to the WHO AQG. Table 1 shows the WHO Air Quality Standards and different limit values for air pollutants for the regions across the world. Figure 1 shows the locations of air quality monitoring stations in Macao. Taipa Ambient is an ambient station in Macao, which is also the background representative station. It is located at Taipa Grande, the headquarter of Macao Meteorological and Geophysical Bureau (SMG).

PM levels are usually measured higher during the winter season, from December to February due to the northern wind bringing the air pollutants to the region, lowering mixing height, fewer amount of rainfall and lower frequency of rainfall. In contrast, the levels of PM are usually measured lower during the summer season, from June to August due to the southern winds from the China sea, higher mixing height and higher frequency of rainfall and amount, which allowed for a better air pollution dispersion and deposition conditions [12], [13]. Figures 2 and 3 show the monthly levels of PM in Macao. The months from June to August recorded the lowest concentration of PM<sub>10</sub> and PM<sub>2.5</sub> in Macao. In contrast, the months from December to February recorded the highest concentration of PM<sub>10</sub> and PM<sub>2.5</sub> in Macao.

Figures 4 and 5 show the hourly levels of PM in Macao. The hours from 10:00 to 12:00 and from 15:00 to 18:00 recorded the highest concentration of PM<sub>10</sub> in Macao. The hours from 8:00 to 11:00 recorded the highest concentration of PM<sub>2.5</sub> in Macao. In addition, Macao has a typical tropical oceanic climate which is hot and humid, with an annual average temperature of 22.3°C and an annual average wind speed of 3.5 m/s, with northwestern wind dominant in winter and southeastern wind dominant in summer [14].

The forecasting of air pollutant concentrations is very important for areas with air quality problems. The forecasting can be developed through the integration of physicochemical relationships from both meteorology and pollutant behavior or by using stochastic methods

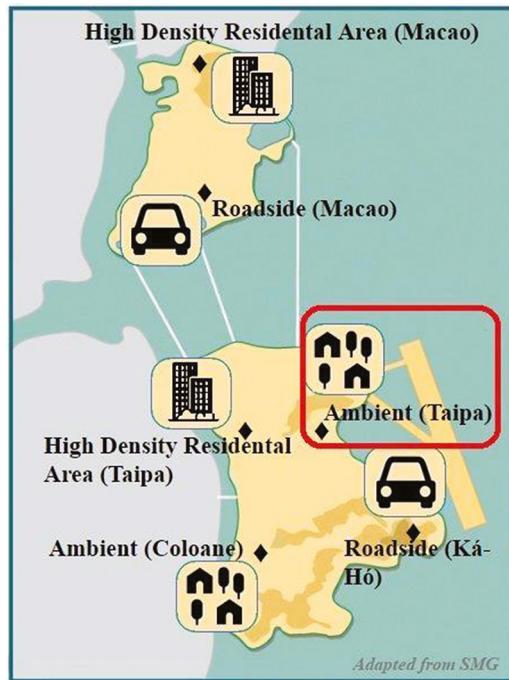


Figure 1: Map of Macao air quality monitoring station network [11].

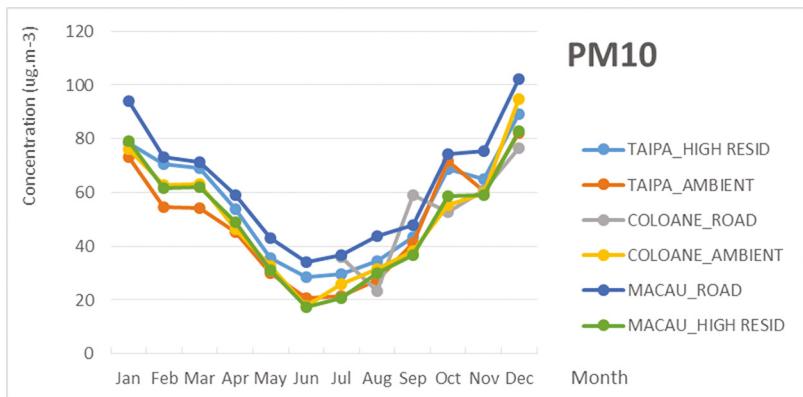


Figure 2: Monthly levels of PM<sub>10</sub> in the Macao air quality monitoring stations from 2013 to 2017.

based on data series analysis. A combination of standard statistical methods was the selected process described in this paper. Statistical models based on multiple regression (MR) analysis were developed to forecast the average daily concentration for PM (PM<sub>10</sub> and PM<sub>2.5</sub>, coarse and fine particles, respectively) for the next day, for the air quality monitoring station of Taipa Ambient.

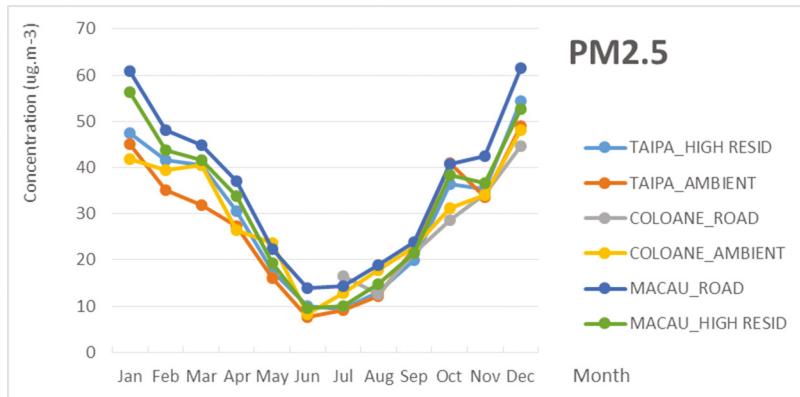


Figure 3: Monthly levels of PM<sub>2.5</sub> in the Macao air quality monitoring stations from 2013 to 2017.

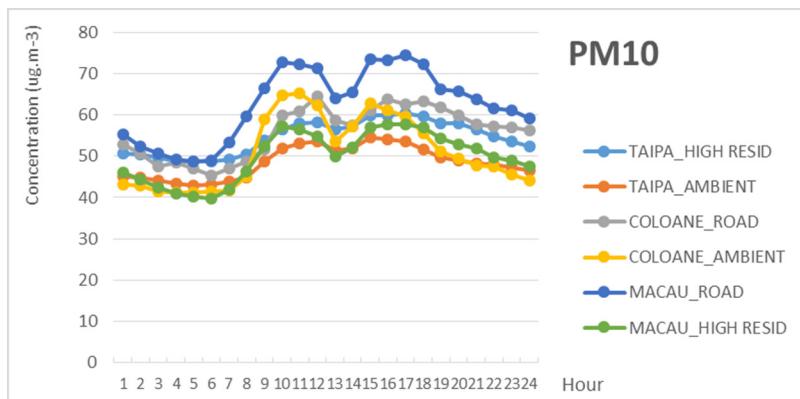


Figure 4: Hourly levels of PM<sub>10</sub> in the Macao air quality monitoring stations from 2013 to 2017.

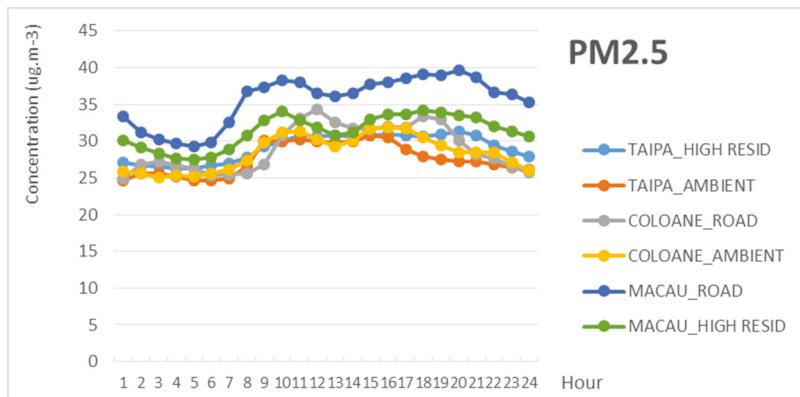


Figure 5: Hourly levels of PM<sub>2.5</sub> in the Macao air quality monitoring stations from 2013 to 2017.

## 2 METHODS

Using past information on studies to understand the variability of  $\text{PM}_{10}$  and  $\text{PM}_{2.5}$  [15]–[19] was the first step to start to build a highly detailed database using all the air quality and meteorological existing data from the years of 2013 to 2017. Data from the years between 2013 and 2016 were used to develop the models, and each of the models was validated using the data of 2017. There were a total of 35 variables to forecast each pollutant, which include 30 meteorological variables and 5 air quality variables. Some of the meteorological variables that were considered included H\_1000, H\_850, H\_700 (the height of geopotential at 1,000, 850 and 700 hPa, respectively), HRMD, HRMN and HRMX (relative humidity of mean, minimum and maximum daily values). Some of the air quality variables that were considered included  $\text{PM}_{10-16D1}$ ,  $\text{PM}_{2.5-16D1}$  ( $\text{PM}_{10}$  and  $\text{PM}_{2.5}$  from the 16:00 of yesterday to 15:00 of today, respectively) and  $\text{PM}_{10-23D1}$  and  $\text{PM}_{2.5-23D1}$  ( $\text{PM}_{10}$  and  $\text{PM}_{2.5}$  average from the previous day, respectively). Table 2 shows the list of meteorological variables that were selected as significant independent variables for the best-fitted MR models, and Table 3 shows the list of air quality variables also incorporated as independent variables in each of the models, for daily next-day average  $\text{PM}_{10}$  and for daily next-day average  $\text{PM}_{2.5}$ , respectively.

Air quality data were obtained from the Macao Meteorological and Geophysical Bureau (SMG). Surface meteorological parameters such as relative humidity were collected from the Taipa Ambient station. Geopotential heights were also collected from the daily soundings of Hong Kong King's Park meteorological station. The European Centre for Medium-Range Weather Forecasting would issue the meteorological forecast at GMT+8 16:00 (local time) daily, and these data are used to feed the forecasting model [20]. In addition, the use of MR can determine the meteorological comparability between the measurement days [21]. Following the precedent experiences [22]–[25], the statistical models were initially created using forward MR analysis with a significance level of 0.05. The MR analysis was performed with SPSS version 25.

Table 2: List and description of meteorological variables to forecast  $\text{PM}_{10}$  and  $\text{PM}_{2.5}$ .

Meteorological variables	Description (units)
H_850	Height of geopotential at 850 hPa (m)
HRMD	Average relative air humidity (%)

Table 3: List and description of air quality variables to forecast  $\text{PM}_{10}$  and  $\text{PM}_{2.5}$ .

Air quality variables	Description (units)
$\text{PM}_{10-16D1}$	Average of the hourly values between 16:00 of yesterday and 15:00 of today for $\text{PM}_{10}$ ( $\mu\text{g}/\text{m}^3$ )
$\text{PM}_{2.5-16D1}$	Average of the hourly values between 16:00 of yesterday and 15:00 of today for $\text{PM}_{2.5}$ ( $\mu\text{g}/\text{m}^3$ )

### 3 RESULTS AND DISCUSSION

Table 4 shows the variables used in each of the statistical models to predict next-day daily average of PM<sub>10</sub> and PM<sub>2.5</sub>. In the case of relative humidity, the decrease in humidity will imply intensification of the northern or western flow with transport of continental dryer air over Macao increasing the concentrations of PM. In the case of the geopotential height at 850 hPa, the influence is less relevant and more difficult to explain because the variability of geopotential height is small at the latitude of Macao. One of the possible explanations for the influence of geopotential height is related to the atmospheric circulation and air mass characteristics of Macao.

Data from 2013 to 2017 were used in this study. In particular, the data from 2013 to 2016 were used to build the statistical models, while the data of 2017 were used to validate the models. The statistical models were built for Taipa Ambient air quality monitoring station of Macao. The corresponding ANOVA tables for the regression models for the selected MR models are presented in Tables 5 and 6 for PM<sub>10</sub> and in Tables 7 and 8 for PM<sub>2.5</sub>, respectively. The equation for next-day 24-h average PM<sub>10</sub> at Taipa Ambient is presented as follows:

$$\text{PM}_{10} = (0.891 \times \text{PM}_{10\_16D1}) + (0.018 \times \text{H}_850) - (0.261 \times \text{HRMD\_TG}) \quad (1)$$

The equation for next-day 24-h average PM<sub>2.5</sub> at Taipa Ambient is presented as follows:

$$\text{PM}_{2.5} = (0.918 \times \text{PM}_{2.5\_16D1}) + (0.009 \times \text{H}_850) - (0.128 \times \text{HRMD\_TG}) \quad (2)$$

Table 4: Variables used in statistical models.

Station	Dependent variables (daily next-day average)	Independent variables used in the best-fit multiple regression model
Taipa Ambient	PM10	PM10_16D1, H_850, HRMD
	PM2.5	PM2.5_16D1, H_850, HRMD

Table 5: ANOVA for PM<sub>10</sub> at Taipa Ambient.

Model		Sum of squares	Df	Mean square	F	Sig.
1	Regression	4,826,913.618	3	1,608,971.206	20,087.037	0.000 <sup>c</sup>
	Residual	112,940.966	1,410	80.100		
	Total	4,939,854.584 <sup>d</sup>	1,413			

Table 6: Model summary for PM<sub>10</sub> at Taipa Ambient.

Model	R	R square <sup>b</sup>	Adjusted R Square	Std. error of the estimate
1	0.989 <sup>a</sup>	0.977	0.977	8.950

Table 7: ANOVA for PM<sub>2.5</sub> at Taipa Ambient.

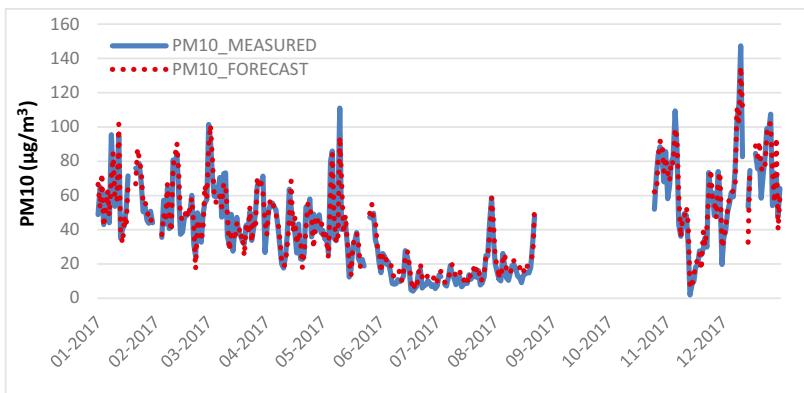
Model		Sum of squares	Df	Mean square	F	Sig.
1	Regression	1,805,650.479	3	601,883.493	14,496.557	0.000 <sup>c</sup>
	Residual	58,085.172	1,399	41.519		
	Total	1,863,735.651 <sup>d</sup>	1,402			

Table 8: Model summary for PM<sub>2.5</sub> at Taipa Ambient.

Model	R	R square <sup>b</sup>	Adjusted R square	Std. error of the estimate
1	0.984 <sup>a</sup>	0.969	0.969	6.444

The models were tested with collected data from 2017 to perform a model validation. The results show a high correlation (with an  $R^2$  of 0.98 and 0.97, for PM<sub>10</sub> and PM<sub>2.5</sub>, respectively), statistically significant at a 95% confidence level. The selected models allow a better understanding of each synoptic situation and its relationship with air quality, leading to a forecast with considerable certainty for the majority of the identified scenarios. Figures 6 and 7 show an example of the test for the year of 2017 and the results of model validation for PM<sub>10</sub> and PM<sub>2.5</sub>.

Table 9 shows the model performance indicators for PM in Taipa Ambient station. The Root Mean Square Error (RMSE) is used to measure the disagreement between the regression model and the observations, and it is particularly sensible to large deviations. Since the BIAS is positive, it means that the model is overestimating on average by 1.1  $\mu\text{g}/\text{m}^3$ . In addition, the Absolute Error (MAE) values of 4.5 and 3.2  $\mu\text{g}/\text{m}^3$  correspond to a relative error around 10% of the mean observed concentrations. The small differences between MAE and RMSE do not indicate the presence of large deviations. The  $R^2$  of around 0.9 means that the predictors explain 90% of variance.

Figure 6: Observed and predicted PM<sub>10</sub> concentration values using MR models in Taipa Ambient (2017).

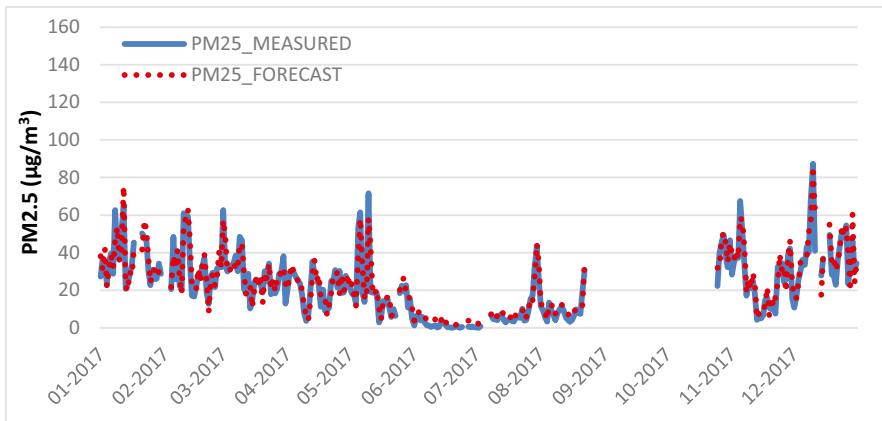


Figure 7: Observed and predicted PM<sub>2.5</sub> concentration values using MR models in Taipa Ambient (2017).

Table 9: Model performance indicators.

Station	Pollutant	Performance indicators			
		R <sup>2</sup>	RMSE ( $\mu\text{g}/\text{m}^3$ )	MAE ( $\mu\text{g}/\text{m}^3$ )	BIAS ( $\mu\text{g}/\text{m}^3$ )
Taipa	PM10	0.92	6.8	4.5	1.1
Ambient	PM2.5	0.89	5.0	3.2	1.1

#### 4 CONCLUSION

The work presented here is a statistical attempt to forecast air pollutant concentrations, based on a detailed analysis of both historical and expert knowledge involving meteorology and air quality aspects concerning PM. The final objective was to develop a daily air quality forecast using statistical methods to predict the next-day daily average of PM<sub>10</sub> and PM<sub>2.5</sub> for the Macao region for the most relevant background location, namely Taipa Ambient. Both models of PM<sub>10</sub> and PM<sub>2.5</sub> used independent variables including the average of the hourly values between 16:00 of yesterday and 15:00 of today for PM<sub>10</sub> and PM<sub>2.5</sub> respectively, geopotential height at 850 hPa and average relative air humidity. A four-year period (2013–2016) was selected as the fitness period for the models, while another period (2017) was selected for the validation of the model. The use of statistical models based on MR analysis was successful in forecasting the average daily concentrations for PM for next day for this particular location in the region of Macao. The models developed also allow for a better understanding of different variables and the relationship between them. The variables that explained most of the variability for PM are geopotential height and average relative air humidity.

#### ACKNOWLEDGMENTS

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

- [1] Sheng, N. & Tang, U.W., Risk assessment of traffic-related air pollution in a world heritage city. *International Journal of Environmental Science and Technology*, **10**(1), pp. 11–18, 2013.
- [2] Statistics and Census Service (DSEC), Macao in Figures. Available at <http://www.dsec.gov.mo/Statistic.aspx?NodeGuid=ba1a4eab-213a-48a3-8fbb-962d15dc6f87>, 2018 (accessed 1 March 2019).
- [3] United States Environmental Protection Agency (USEPA), Particle Pollution and Your Health. Available at <https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=P1001EX6.txt>, (accessed 1 March 2019).
- [4] World Health Organization (WHO), Health Aspects of Air Pollution with Particulate Matter, Ozone and Nitrogen Dioxide. Available at [http://www.euro.who.int/\\_data/assets/pdf\\_file/0005/112199/E79097.pdf](http://www.euro.who.int/_data/assets/pdf_file/0005/112199/E79097.pdf), (accessed 1 March 2019).
- [5] Ministry of Ecology and Environment (MEE), Ambient Air Quality Standards. Available at <http://210.72.1.216:8080/gzaqi/Document/gjzlbz.pdf>, (accessed 1 March 2019).
- [6] Krzyzanowski, M. & Cohen, A., Update of WHO air quality guidelines. pp. 7–13, 2008.
- [7] World Health Organization (WHO), Air Quality Guidelines. Available at [http://202.171.253.71/www.euro.who.int/\\_data/assets/pdf\\_file/0005/78638/E90038.pdf](http://202.171.253.71/www.euro.who.int/_data/assets/pdf_file/0005/78638/E90038.pdf), (accessed 1 March 2019).
- [8] World Health Organization (WHO), WHO Air Quality Guidelines for Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide. Available at [http://202.171.253.72/apps.who.int/iris/bitstream/handle/10665/69477/WHO\\_SDE\\_PHE\\_OEH\\_06.02\\_eng.pdf?sequence=1&isAllowed=y](http://202.171.253.72/apps.who.int/iris/bitstream/handle/10665/69477/WHO_SDE_PHE_OEH_06.02_eng.pdf?sequence=1&isAllowed=y), (accessed 1 March 2019).
- [9] World Health Organization (WHO), Evolution of WHO Air Quality Guidelines: Past, Present and Future. Available at [http://202.171.253.69/www.euro.who.int/\\_data/assets/pdf\\_file/0019/331660/Evolution-air-quality.pdf](http://202.171.253.69/www.euro.who.int/_data/assets/pdf_file/0019/331660/Evolution-air-quality.pdf), (accessed 1 March 2019).
- [10] World Health Organization (WHO), WHO Expert Consultation : Available Evidence for the Future Update of the WHO Global Air Quality Guidelines (AQG). Available at [http://202.171.253.66/www.euro.who.int/\\_data/assets/pdf\\_file/0013/301720/Evidence-future-update-AQGs-mtg-report-Bonn-sept-oct-15.pdf](http://202.171.253.66/www.euro.who.int/_data/assets/pdf_file/0013/301720/Evidence-future-update-AQGs-mtg-report-Bonn-sept-oct-15.pdf), (accessed 1 March 2019).
- [11] Macao Meteorological and Geophysical Bureau (SMG), Resumo anual sobre qualidade do ar em Macau – 2017. Available at [http://www.smg.gov.mo/smgs/airQuality/pdf/IQA\\_2017\\_PT.pdf](http://www.smg.gov.mo/smgs/airQuality/pdf/IQA_2017_PT.pdf), (accessed 1 March 2019).
- [12] Lopes, D., Ferreira, J., Hoi, K.I., Miranda, A.I., Yuen, K.V. & Mok, K.M., Weather research and forecasting model simulations over the Pearl River Delta Region. *Air Quality, Atmosphere and Health*, pp. 115–125, 2018.
- [13] Lopes, D., Hoi, K.I., Mok, K.M., Miranda, A.I., Yuen, K.V. & Borrego, C., Air quality in the main cities of the Pearl River Delta Region. *Global Nest Journal*, **18**(4), pp. 794–802, 2016.
- [14] He, D., Zhou, Z., He, K., Hao, J., Liu, Y., Wang, Z. & Deng, Y., Assessment of traffic related air pollution in urban areas of Macao. *Journal of Environmental Sciences*, **12**(1), pp. 39–46, 2000.
- [15] Ferreira, F.C., Torres, P.M., Tente, H.S. & Neto, J.B., Ozone levels in Portugal: the Lisbon region assessment, 2004.
- [16] Clapp, L.J. & Jenkin, M.E., Analysis of the relationship between ambient levels of O<sub>3</sub>, NO<sub>2</sub> and NO as a function of NO<sub>x</sub> in the UK. *Atmospheric Environment*, **35**(36), pp. 6391–6405, 2001.

- [17] Ferreira, F., Tente, H., Torres, P., Cardoso, S. & Palma-Oliveira, J., Air quality monitoring and management in Lisbon. *Environmental Monitoring and Assessment*, **65**, pp. 443–450, 2000.
- [18] Neto, J., Ferreira, F., Torres, P.M. & Boavida, F., Lisbon air quality forecast using statistical methods. *International Journal of Environment Pollution*, **39(3)**, pp. 333–340, 2009.
- [19] Wang, W., Lu, W., Wang, X. & Leung A.Y.T., Prediction of maximum daily ozone level using combined neural network and statistical characteristics. *Environment International*, **29(5)**, pp. 555–562, 2003.
- [20] European Centre for Medium-Range Weather Forecasts (ECMWF), User Guide to ECMWF Forecast Products, Version 4.0. Available at [https://www.uio.no/studier/emner/matnat/geofag/nedlagte-emner/GEF4220/v09/undervisningsmateriale/Persson\\_user\\_guide.pdf](https://www.uio.no/studier/emner/matnat/geofag/nedlagte-emner/GEF4220/v09/undervisningsmateriale/Persson_user_guide.pdf), (accessed 1 March 2019).
- [21] Choi, W., Paulson, S.E., Casmassi, J. & Winer, A.M., Evaluating meteorological comparability in air quality studies: classification and regression trees for primary pollutants in California's South Coast Air Basin. *Atmospheric Environment*, **64**, pp. 150–159, 2013.
- [22] United States Environmental Protection Agency (USEPA), Guidelines for Developing an Air Quality (Ozone and PM2.5) Forecasting Program. Available at [http://infohawk.uiowa.edu/F/19FJHSI1VLLSQ1YJQLQ7SK1BIG4QCC5N92I2J7Y3LIFRT6CDFN-03298?func=full-set-set&set\\_number=002362&set\\_entry=000004&format=999](http://infohawk.uiowa.edu/F/19FJHSI1VLLSQ1YJQLQ7SK1BIG4QCC5N92I2J7Y3LIFRT6CDFN-03298?func=full-set-set&set_number=002362&set_entry=000004&format=999), (accessed 1 March 2019).
- [23] Cassmassi, J.C., 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, 1997.
- [24] Oduro, S.D., Ha, Q.P. & Duc, H., Vehicular emissions prediction with CART-BMARS hybrid models. *Transportation Research Part D: Transport and Environment*, **49**, pp. 188–202, 2016.
- [25] Durão, R.M., Mendes, M.T. & Pereira, M.J., Forecasting O<sub>3</sub> levels in industrial area surroundings up to 24 h in advance, combining classification trees and MLP models. *Atmospheric Pollution Research*, **7**, pp. 961–970, 2016.