



# Air Quality Forecast by Statistical Methods: Application to Portugal and Macao

Lúisa Mendes<sup>1\*</sup>, Joana Monjardino<sup>1,2</sup> and Francisco Ferreira<sup>1,2</sup>

<sup>1</sup> Department of Environmental Sciences and Engineering, NOVA School of Sciences and Technology, NOVA University Lisbon, Lisbon, Portugal, <sup>2</sup> Center for Environmental and Sustainability Research, NOVA School of Sciences and Technology, NOVA University Lisbon, Lisbon, Portugal

## OPEN ACCESS

### Edited by:

Rasa Zalakeviciute,  
University of the Americas, Ecuador

### Reviewed by:

Steffen M. Noe,  
Estonian University of Life  
Sciences, Estonia  
Qutu Jiang,  
The University of Hong Kong, Hong  
Kong SAR, China

### \*Correspondence:

Lúisa Mendes  
lc.mendes@fct.unl.pt

### Specialty section:

This article was submitted to  
Data-driven Climate Sciences,  
a section of the journal  
Frontiers in Big Data

Received: 30 November 2021

Accepted: 08 February 2022

Published: 10 March 2022

### Citation:

Mendes L, Monjardino J and  
Ferreira F (2022) Air Quality Forecast  
by Statistical Methods: Application to  
Portugal and Macao.  
Front. Big Data 5:826517.  
doi: 10.3389/fdata.2022.826517

Air pollution is a major concern issue for most countries in the world. In Portugal and Macao, the values of nitrogen dioxide (NO<sub>2</sub>), particulate matter (PM) and ozone (O<sub>3</sub>) are frequently above the concentration thresholds accepted as “good air quality.” Portugal follows the European Union (EU) legislation (Directive 2008/50/EC) on air quality and Macao the air quality guidelines (AQG) from the WHO. Air quality forecasts are very important mitigation tools because of their ability to anticipate pollution events, and issue early warnings, allowing to take preventive measures and reduce impacts, by avoiding exposure. The work presented here refers to the statistical forecast of air pollutants for three regions: Greater Lisbon Area, Madeira Autonomous Region (both located in Portugal), and Macao Special Administrative Region (in Southern China). The presented statistical approach combines Classification and Regression Tree (CART) and multiple regression (MR) analysis to obtain optimized regression models. This consolidated methodology is now in operation for more than a decade in Portugal, and is subject to regular updates that reflect the ongoing research and the changes in the air quality monitoring network. Recently, the same methodology was applied to Macao in collaboration with the Macao Meteorological and Geophysical Bureau (SMG). Here, a statistical approach for air quality forecasting is described that has been proven to be successful, being able to forecast PM<sub>10</sub>, PM<sub>2.5</sub>, NO<sub>2</sub>, and O<sub>3</sub> concentrations, for the next day, with a good performance. In general, all the models have shown a good agreement between the observed and forecasted concentrations (with  $R^2$  from 0.50 to 0.89), and were able to follow the concentration evolution trend. For some cases, there is a slight delay in the prediction trend. Moreover, the results obtained for pollution episodes have proven that statistical forecast can be an effective way of protecting public health.

**Keywords:** particulate matter, ozone, nitrogen dioxide, air quality, classification and regression trees, multiple regression

## INTRODUCTION

The Ambient Air Quality Directives of European Union (EU) set standards for key air pollutants. These values take into account the 2005 WHO guidelines and considerations of technical and economic feasibility at the time of their adoption.

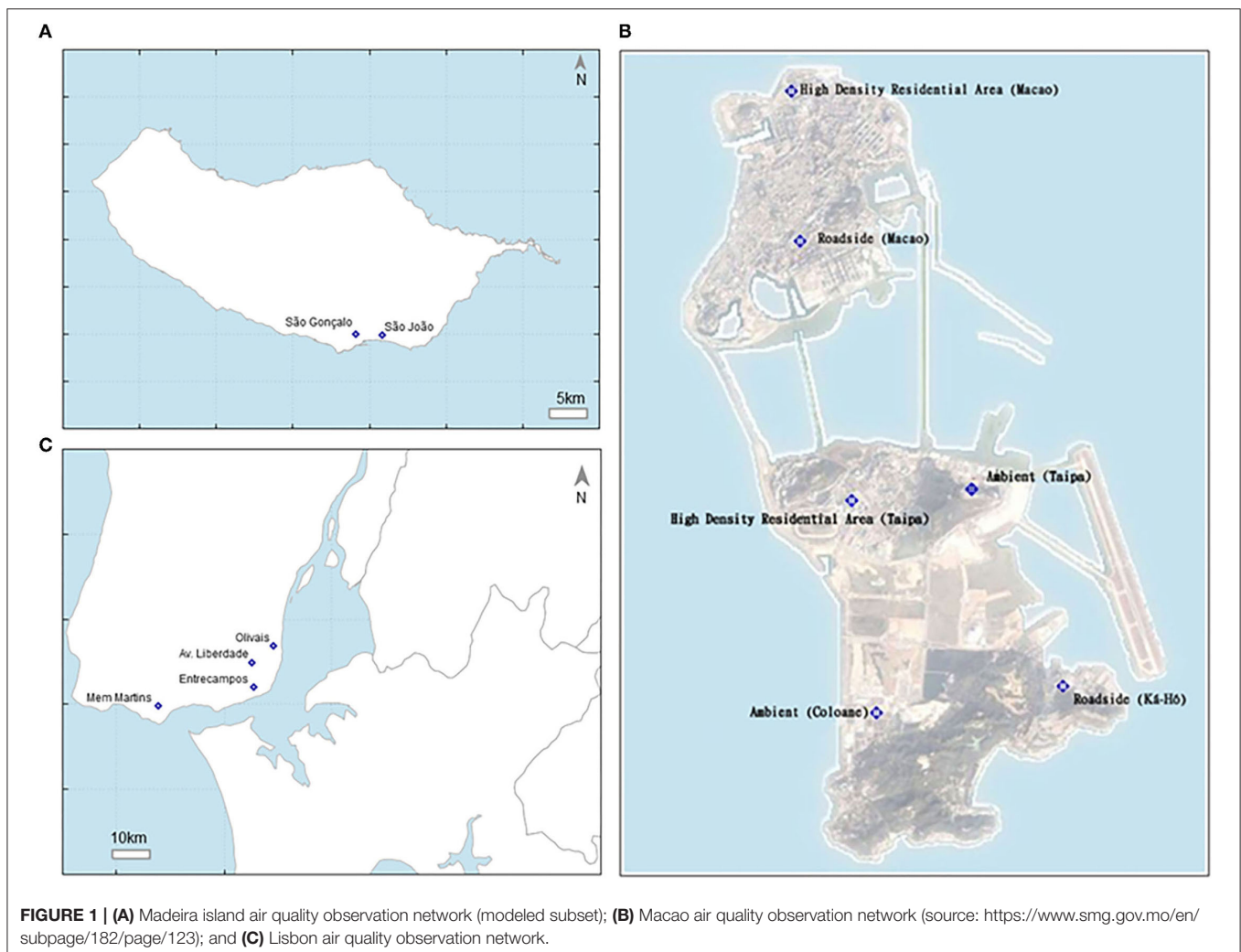
Air quality forecasting, if reliable and sufficiently accurate, can play an important role as part of an air quality management system (NOAA, 2001). Its applications can fall into several broad areas,

such as health alerts—many cities currently provide warnings to the public when air pollution levels exceed specified levels, being those warnings directed at specific populations that are particularly sensitive to air pollution (e.g., asthmatics) (Liu et al., 2018); in addition, air quality forecasts can supplement existing emission control programs or emergency responses, with cities offering free access to public transportation (Quarmby et al., 2019); on pollution episode days, to reduce vehicle emissions, and regions implementing the “No-Burn day” (AQMD, 2022); consisting of a ban period on wood-burning in residential fireplaces, stoves, or outdoor fire pits, when particulate matter concentrations are expected to reach unhealthy levels, due to air emissions and stagnant weather conditions.

To predict the next-day daily average concentrations of particulate matter ( $PM_{10}$  and  $PM_{2.5}$ ), daily hourly maximum concentrations of ozone ( $O_3$ ), and daily hourly maximum concentrations of nitrogen dioxide ( $NO_2$ ), at air quality monitoring stations locations, forecast models were developed based on statistical methods using multiple linear regression (MR) and Classification and Regression Tree (CART) analysis.

The NOVA University Lisbon (NOVA School of Science and Technology), in collaboration with the Portuguese Environment Agency (APA) and the Portuguese Institute for Sea and Atmosphere (IPMA), runs and disseminates daily air quality forecasts based on a statistical approach, first used by Cassmassi (1987) at South Coast Air Quality Management District California, USA. This statistical methodology is now in operation, in Portugal (Neto et al., 2005), for more than a decade and is the subject of regular updates, reflecting the ongoing research, and the changes in the air quality monitoring network. Recently the same methodology was extended to Madeira Autonomous Region, in Portugal, and was also applied to Macao Special Administrative Region of the People’s Republic of China (MSAR), resulting from a collaboration with the Macao Meteorological and Geophysical Bureau (SMG) (Lei et al., 2019, 2020).

Air pollution is a major concern issue for most countries in the world. The global burden of disease associated with air pollution exposure exacts massive toll on human health worldwide: it is estimated to cause millions of deaths and lost years of healthy



life annually. The burden of disease attributable to air pollution is now estimated to be on a par with other major global health risks, such as unhealthy diet and tobacco smoking, and air pollution is now recognized as the single biggest environmental threat to human health (World Health Organization, 2021). In Portugal, despite improvements in the past two decades, there are still exceedances mainly to nitrogen dioxide (NO<sub>2</sub>) annual limit value, to particulate matter (PM<sub>10</sub>) daily limit value, and to ozone (O<sub>3</sub>) target value. In Macao, concentrations of these pollutants are frequently above the thresholds accepted as “good air quality.”

High concentrations of NO<sub>2</sub>, PM, and O<sub>3</sub> in the low troposphere are an additional risk factor for cardiovascular and respiratory diseases and contribute to mortality all over the world (Sheng and Tang, 2013; Lee et al., 2017). Surface O<sub>3</sub> is known by its negative impacts in the respiratory system leading to more hospitalizations (Entwistle et al., 2019). PM, in particular smaller fractions such as PM<sub>2.5</sub>, are a major concern once they can get deep into lungs and some may even get into the bloodstream (Wiśniewska et al., 2019). Finally, specific combinations of concentration levels of these pollutants may be more dangerous than equally high levels of all the pollutants (Sheng and Tang, 2013). Portugal follows the European Union (EU) legislation (Directive 2008/50/EC) (European Union Legislation, 2008) on air quality, and Macao follows the Chinese National Ambient Air Quality Standards (NAAQS), which in turn are based in the Interim target-1 Air Quality Guidelines from the WHO (WHO, 2006). EU limit values, legally binding, were set for human health protection for, among other pollutants and averaging periods, PM<sub>10</sub> as 50 µg/m<sup>3</sup> (at a daily basis), PM<sub>2.5</sub> as 25 µg/m<sup>3</sup> (at annual basis), NO<sub>2</sub> as 40 µg/m<sup>3</sup> (at annual basis), and the O<sub>3</sub> target value is of 120 µg/m<sup>3</sup> (referring to the maximum daily 8-h mean, represented as O<sub>3MAX</sub>). The NAAQS has set the threshold of PM<sub>10</sub> at 150 µg/m<sup>3</sup> (daily basis), PM<sub>2.5</sub> at 75 µg/m<sup>3</sup> (daily basis), NO<sub>2</sub> at 40 µg/m<sup>3</sup> (annual basis), and O<sub>3MAX</sub> at 160 µg/m<sup>3</sup> (daily maximum 8-h mean). Some of these values contrast with the recently recommended WHO Air Quality Guideline levels, updated in 2021, after a systematic review of accumulated evidence, which are set at 45 µg/m<sup>3</sup> for PM<sub>10</sub> (daily basis), 15 µg/m<sup>3</sup> (daily basis) and 5 µg/m<sup>3</sup> (annual basis) for PM<sub>2.5</sub>, 10 µg/m<sup>3</sup> for NO<sub>2</sub> (annual basis), and 100 µg/m<sup>3</sup> for O<sub>3</sub> (daily maximum 8-h mean) (World Health Organization, 2021).

High-density and high-rise cities have become increasingly common in Asia (Lee et al., 2017). Air quality is a significant public health risk in many of these cities (World Health Organization, 2016). Macao, a coastal city located in southern China, is one example of a high-density, high-rise city with air quality issues. Macao was listed as the number one most densely populated region in the world (Sheng and Tang, 2013), with a population density of about 20,000 inhabitants/km<sup>2</sup>, accounting for a population of 680,000 within an area of 32.9 km<sup>2</sup>. The clustering effect is further enhanced by the prevalence of high-rise buildings. European cities likely have lower pollution and building densities, and fewer small-scale dispersed pollution sources, than high-density high-rise cities.

Factors leading to variation in pollution levels are diverse, and include both human activities and meteorological factors (Boubel

**TABLE 1** | Modeling and validation periods considered by region and air pollutants forecasted at each air quality monitoring station (AQMS).

Region	Modeling period	Validation period	AQMS (Type)	Modeled Pollutants
Greater Lisbon Area	2015–2018	2019	Av. Liberdade (UT)	PM <sub>10</sub> , NO <sub>2</sub>
			Entrecampos (UT)	PM <sub>10</sub> , PM <sub>2.5</sub> , NO <sub>2</sub>
			Olivais (UB)	PM <sub>10</sub> , PM <sub>2.5</sub> , NO <sub>2</sub> , O <sub>3</sub>
Macao administrative region	2013–2018	2019	São João (UT)	PM <sub>10</sub> , PM <sub>2.5</sub> , NO <sub>2</sub>
			São Gonçalo (UB)	PM <sub>10</sub> , O <sub>3</sub>
Macao autonomous region	2016–2018	2019	Macao Roadside (UT)	PM <sub>10</sub> , PM <sub>2.5</sub> , NO <sub>2</sub>
			Macao Residential (HDR)	PM <sub>10</sub> , PM <sub>2.5</sub> , NO <sub>2</sub> , O <sub>3</sub>
			Taipa Ambient (UB)	PM <sub>10</sub> , PM <sub>2.5</sub> , NO <sub>2</sub> , O <sub>3</sub>
Macao administrative region	2013–2018	2019	Taipa Residential (HDR)	PM <sub>10</sub> , PM <sub>2.5</sub> , NO <sub>2</sub> , O <sub>3</sub>

AQMS, Air Quality Monitoring Station; UT, Urban Traffic; UB, Urban Background; HDR, High Density Residential.

et al., 1994). Meteorology plays a fundamental role in the re-distribution of air pollutants after their release in the atmosphere (Boubel et al., 1994). The characterization of local and large scale circulation winds and vertical atmospheric stability, allows accounting for transport, mixture, and dispersion processes (Boubel et al., 1994). Precipitation refers to the natural processes by which material is removed by atmospheric hydrometeors (cloud and fog drops, rain, and snow) and delivered to the Earth's surface (Seinfeld and Pandis, 2006).

The association between specific meteorological parameters and air quality can be quantified using a variety of statistical techniques. Statistical forecast methods analyze the events without knowing the mechanism of the change; therefore, this method is not dependent on physical, chemical, or biological processes (Bai et al., 2018). Instead, methods, such as regression analysis, investigate relationships between variables. Forecasting is a requisite part in the science of big data, and can be used to infer the future development of an object relative to previous information (Bai et al., 2018). Pollution forecasting can be understood as an estimation of a pollutant concentration at a specified future date.

This work aims to provide an overall description of the current statistical methods, used by NOVA University Lisbon air quality group, to forecast air pollutant concentrations. Some of the discussed aspects are related with data requirements, steps involved in the model development, advantages and disadvantages of this approach. Model performance indicators are presented for each region and pollutant. Finally, examples of model performance are presented, for pollution episodes

**TABLE 2** | Data sources and variables on a daily temporal scale used in the modeling process.

Data type	Source	Variables	Description
Meteorological data	Upper air meteorological observations (Aerological soundings)	H_1000, H_850, H_700, H_500	Geopotential height at pressure levels (indicator of synoptic-scale weather pattern) (hPa)
		TAIR_925, TAIR_850 and TAIR_700	Air temperature at pressure levels (measure of the strength and height of subsidence inversion) (°C)
		RH_925, RH_850, RH_700	Relative humidity at pressure levels (%)
		DEWP_925, DEWP_850, DEWP_700	Dew point at pressure levels (°C)
		THI_850, THI_700, THI_500	Thickness at pressure levels (associated to the mean temperature in the layer)
	STB_925, STB_850, STB_700	Stability at pressure levels (detector of atmospheric stability)	
	Surface meteorological observations (hourly data)	TAIR <sub>MEA</sub> , TAIR <sub>MIN</sub> , TAIR <sub>MAX</sub>	Air temperature, mean, minimum and maximum (air stability and emission rates from engines) (°C)
		RH <sub>MEA</sub> , RH <sub>MAX</sub> , RH <sub>MIN</sub>	Relative humidity, daily mean, maximum and minimum values (%)
		DEWP <sub>MEA</sub>	Mean dew point (°C)
		V <sub>MEA</sub> , V <sub>MAX</sub>	Wind speed mean and maximum values (horizontal dispersion) (m/s)
PREC		24 h Accumulated Precipitation (pollutant removal indicator) (mm)	
	STA1_P-STA2_P	Pressure difference between stations (indicator of synoptic scale weather) (hPa)	
Air quality data	Surface air quality stations (hourly data)	PM <sub>10</sub> _D1, PM <sub>10</sub> _D2, PM <sub>10</sub> _D3, PM <sub>10</sub> _D12, PM <sub>2.5</sub> _D1, PM <sub>2.5</sub> _D2, PM <sub>2.5</sub> _D3, PM <sub>2.5</sub> _D12	Daily mean concentrations for particulate matter (PM <sub>10</sub> and PM <sub>2.5</sub> ) for the recent past (last 3 days—D1 to D3) and the last 24h from each days noon (D12) (μg/m <sup>3</sup> )
		O <sub>3</sub> _D1, O <sub>3</sub> _D2, O <sub>3</sub> _D3, O <sub>3</sub> _D12, NO <sub>2</sub> _D1, NO <sub>2</sub> _D2, NO <sub>2</sub> _D3, NO <sub>2</sub> _D12	Daily maximum concentrations for ozone (O <sub>3</sub> ) and nitrogen dioxide (NO <sub>2</sub> ) for the recent past (last 3 days—D1 to D3) and the last 24h from each days noon (D12) (μg/m <sup>3</sup> )
Other data	Geographical data and human behavior descriptors	Daylight	Number of hours of daylight (h)
		WW	Week/Weekend indicator flag (human activity and traffic)

occurred in 2019 over the three studied regions. In this context, air quality forecast models are relevant tools because of their ability to anticipate and follow pollution episodes, allowing to support decisions, such as early warnings to the population, which can take preventive measures and avoid exposure, and reducing negative health impacts.

## METHODS

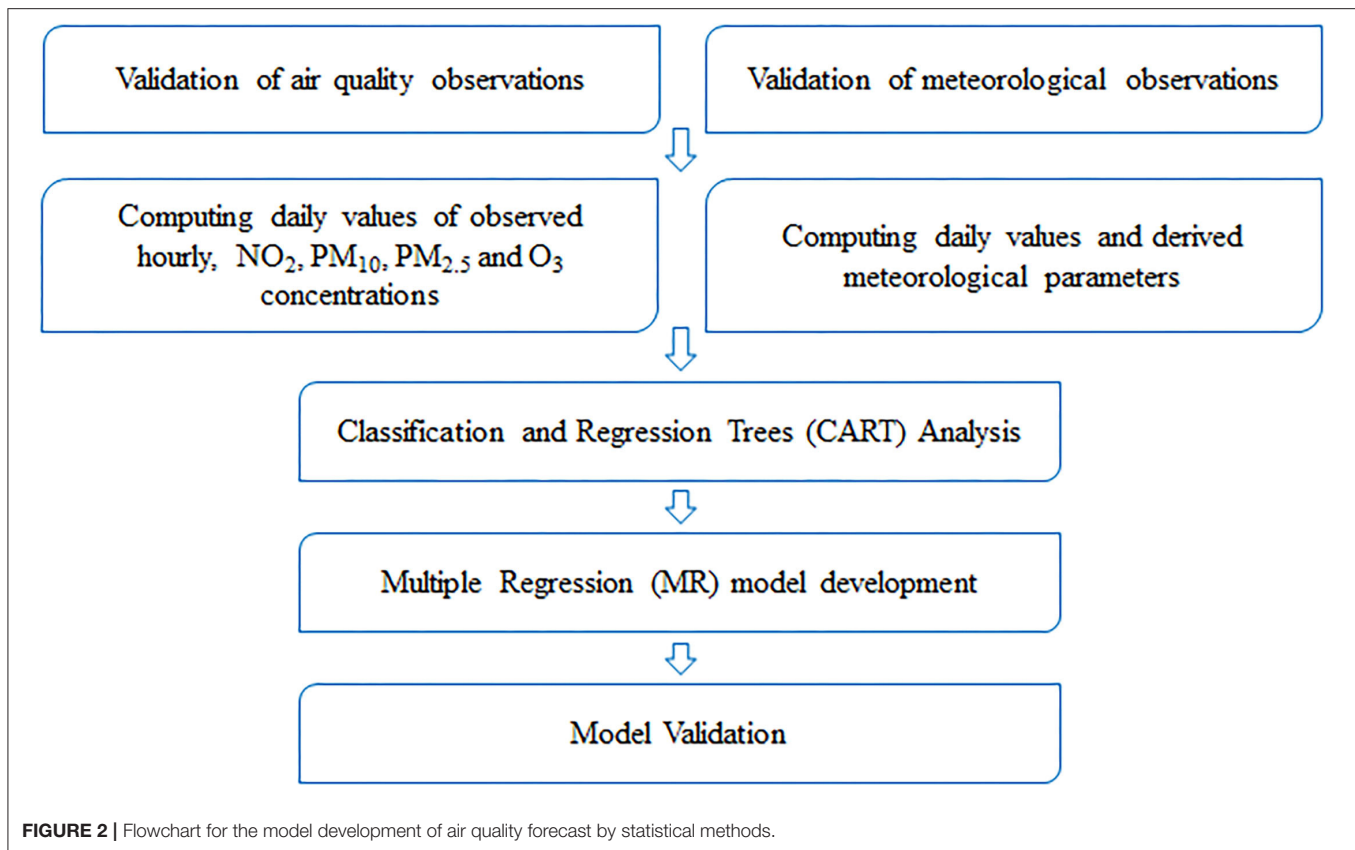
Nowadays, forecasting, by statistical methods, has a wide range of applications and is used all over the world, based on the application of a multitude of algorithms. These methods are very accurate and enable a better understanding of the relationships between air quality data behavior and the underlying meteorology (Bai et al., 2018). In the present work, statistical models were developed based on the techniques of MR and CART. Both techniques rely on the historical data of meteorological and air quality variables.

Regression analysis methods are based on the association between pollutant levels, and meteorological and aerometric variables, which can be quantified by analyzing historical datasets, using standard statistical analysis packages, as shown

on previous works (Cassmassi, 1987; US EPA, 2003; Choi et al., 2013; Durão et al., 2016; Oduro et al., 2016). The resultant multivariate linear regression equation is then used to forecast future pollution levels. The CART technique identifies those variables (meteorological or air quality) that are most strongly correlated with ambient pollution levels. These variables are then used to predict next day pollution levels, either daily average or maximum hourly concentrations depending on the pollutant, based on same day air quality levels and next day forecasted meteorology.

The referred statistical models were applied to forecast the average daily concentrations of PM<sub>10</sub> and PM<sub>2.5</sub>, and the hourly maximum concentrations of O<sub>3</sub> and NO<sub>2</sub> (referred as O<sub>3</sub><sub>MAX</sub> and NO<sub>2</sub><sub>MAX</sub>, respectively), for the next day, for each air quality monitoring station (AQMS) location. For the work presented here, a set of AQMS were selected in the Greater Lisbon Area (4 AQMS), Madeira Autonomous Region (2 AQMS), and Macao (4 AQMS), represented in **Figure 1**.

The Greater Lisbon Area includes the Portuguese capital, Lisbon, being the main economical sub region of the country. It covers 1,376 km<sup>2</sup> and it is the most densely populated Portuguese sub region (with 2,042,477 inhabitants and 1,484



inhabitants/km<sup>2</sup>). The Greater Lisbon Area presents a mild subtropical or a hot-summer Mediterranean climate (Csa), according to Köppen's classification (Köppen, 1936), based on temperature and precipitation mean values. Lisbon summer is mild to hot with significant temperature variations related, in part, to coastal distance. The wind blows most frequently from the north quadrant, according to the 1971–2000 climatological normals (IPMA, 2022).

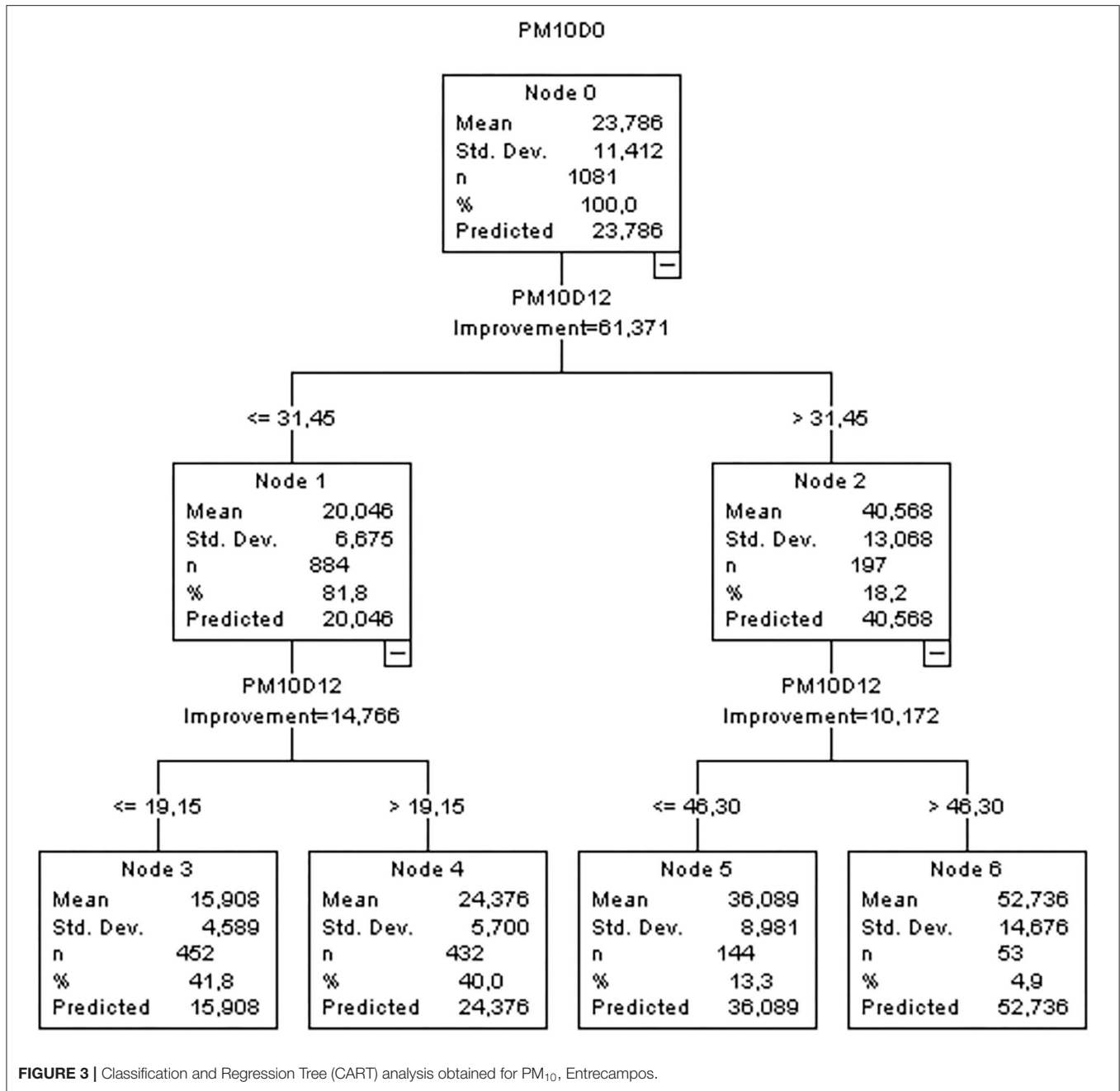
Madeira, an autonomous region of Portugal, is an archipelago comprising four islands off the northwest coast of Africa. This region comprises a population of 253,945 inhabitants within an area of 801 km<sup>2</sup>. The capital of Madeira is Funchal that has a subtropical Mediterranean climate (Csa), according to Köppen's classification (Köppen, 1936). Funchal's climate is predominantly determined by the Atlantic Ocean and as a result weather extremes are rare. Temperature usually rises significantly when the influence of the persistent eastern-winds from northern Africa is felt. The annual average maximum temperature is 22.1°C and the average of minimum temperature is 15.8°C (IPMA, 2022).

Located along the southeast coast of Mainland China, Macao is surrounded by water on three sides, with a subtropical oceanic monsoon climate that is characterized by high temperatures, high levels of atmospheric moisture and abundant rainfall (SMG, 2022). In winter, Macao is cold and dry with predominant northern winds, and the summer is presented with heavy rains due to the strong southwest monsoon. Spring and autumn are

transition periods. The winter northeast monsoon is known to have the ability to transport pollutants from northern and eastern China (Tong et al., 2018). In summer season, from June to August, rainfall increases, providing a better atmospheric mixing, and persistent southern winds occur, resulting in PM levels to decrease (Lopes et al., 2016).

Data from 3- to 6-year daily series observations, were used to develop the forecast models, and each of the models was evaluated using 2019 data. The time period used to build each model equation is different for each region and AQMS, according to data availability (Table 1). The selection of a representative modeling period is important, being recommended at least 2 years of data.

Regarding the data collection phase, a large set of meteorological and air quality data was gathered, namely: (i) meteorological surface observations: hourly observations from automatic weather stations, such as temperature, relative humidity, and dew point temperature; (ii) upper-air observations, such as, geopotential heights, temperature, relative humidity, and dew point temperature at various altitudes; and (iii) surface air quality measurements, from AQMS network, of PM<sub>10</sub>, PM<sub>2.5</sub>, NO<sub>2</sub>, and O<sub>3</sub>. Other variables were added to the analysis, as the flag for week/weekend day and the daily sunlight period duration. A list of independent variables and data sources used as potential predictors in the modeling phase is presented on Table 2. The model development flowchart is represented in Figure 2.



**FIGURE 3 |** Classification and Regression Tree (CART) analysis obtained for PM<sub>10</sub>, Entrecampos.

Series with a low annual efficiency (<75%) on data availability were rejected. For operational purposes, limitations related to the expected daily data availability, to perform the forecast, were also considered. Outliers were identified and excluded from the data series.

In the process of the MR model development, several of the initial variables were removed of the final models due to high correlation values among variables, or due to the commitment to obtain the simplest model, with the smallest number of variables that maximizes the explained variance. In addition, fewer variables mean that, in operational mode, missing data have

a lower impact on the quality of the forecast. In MR analysis, one has to seek for a compromise between model improvement, obtained by adding variables, and the increase of complexity and uncertainty introduced by a new variable.

Another important aspect is the effort to achieve accurate forecast results when higher concentrations are predicted, since they are related with higher negative health impacts. One of the advantages of the CART technique is to be able to establish particular model equations that can accommodate specific trends caused by meteorological circumstances that trigger high level concentrations (Choi et al., 2013). This is important to

**TABLE 3** | Forecast model selected variables for each pollutant and air quality monitoring station.

Location/AQMS/pollutant		Model selected weather variables																	
		H			TAIR 925	RH 925	DEWP 700	THI 700	STB			TAIR		RH		DEWP MEA	V-MAX	PREC	
		1,000	850	500					925	850	700	MAX	MIN	MEA	MIN				
Greater Lisbon Area	ENT	PM <sub>10</sub>	x			x												x	
	AVL		x		x													x	
	OLI		x		x							x							
	MEM		x	x	x														
	ENT	PM <sub>2.5</sub>	x																
	OLI				x														x
	MEM		x		x	x						x							
	ENT	NO <sub>2</sub>	x													x	x		
	AVL		x	x						x				x					
	OLI		x	x										x					
	MEM		x	x														x	x
	ENT	O <sub>3</sub>						x			x	x						x	
	OLI				x			x			x	x	x						
	MEM									x								x	
Madeira Administrative Region	SJO	PM <sub>10</sub>				x											x		
	SGO					x											x		
	SJO	PM <sub>2.5</sub>				x													
	SJO	NO <sub>2</sub>			x	x											x	x	
	SGO	O <sub>3</sub>			x												x		
Macao Autonomous Region	M_RES	PM <sub>10</sub>	x														x		
	M_ROA		x														x		
	T_AMB		x														x		
	T_RES		x														x		
	M_RES	PM <sub>2.5</sub>	x															x	
	M_ROA		x															x	
	T_AMB		x															x	
	T_RES					x												x	
	T_RES	NO <sub>2</sub>	x															x	
	M_ROA		x															x	
	T_AMB		x							x									
	T_RES		x															x	
	M_RES	O <sub>3</sub>																x	
	T_AMB																	x	
	T_RES		x															x	

ENT, Entrecampos; AVL, Avenida da Liberdade; OLI, Olivais; MEM, Mem Martins; SJO, São João; SGO, São Gonçalo; M\_RES, Macao Residential; M\_ROA, Macao Roadside; T\_AMB, Taipa Ambient; T\_RES, Taipa Residential; H, Geopotential height at different pressure levels; TAIR, Air temperature 925 hPa; RH, Relative humidity at 925 hPa; DEWP, Dew Point at 700 hPa; THI, Thickness at 700 hPa; STB, Stability at different pressure levels; TAIR, Temperature, maximum and minimum; RH, Relative humidity, mean and minimum; DEWP<sub>MEA</sub>, Mean dew point; V\_MAX, Maximum wind speed; PREC, Precipitation.

**TABLE 4 |** Model equations obtained for Greater Lisbon Area (Entrecampos), Madeira Autonomous Region (São João), and Macao Administrative Region (Taipa Ambient).

Station	Pollutant	Model equations
Greater Lisbon Area: Entrecampos	NO <sub>2MAX</sub>	<b>CART nodes and multiple regression equations:</b>
		<b>For NO<sub>2</sub>D12 ≤ 75.45</b> NO <sub>2MAX</sub> = NO <sub>2</sub> D12 * 0.784 – (STA1_P – STA2_P) * 5.201 + H_850 * 0.015
		<b>For 75.45 &lt; NO<sub>2</sub>D12 ≤ 120.05</b> NO <sub>2MAX</sub> = NO <sub>2</sub> D12 * 0.424 + H_850 * 0.048 – V <sub>MAX</sub> * 2.169
	PM <sub>10</sub>	<b>For NO<sub>2</sub>D12 &gt; 120.05</b> NO <sub>2MAX</sub> = NO <sub>2</sub> D12 * 0.311 + H_850 * 0.068 – TAIR <sub>MEA</sub> 2.303
		<b>CART nodes and multiple regression equations:</b>
		<b>For PM<sub>10</sub>D12 ≤ 31.45</b> PM <sub>10</sub> = PM <sub>10</sub> D12 * 0.787 + H_850 * 0.003 – (STA1_P – STA2_P) * 0.717 + TAIR_925 * 0.122
	PM <sub>2.5</sub>	<b>For 31.45 &lt; PM<sub>10</sub>D12 ≤ 46.30</b> PM <sub>10</sub> = PM <sub>10</sub> D12 * 0.965 – PREC * 0.421
		<b>For PM<sub>10</sub>D12 &gt; 46.30</b> PM <sub>10</sub> = H_850 * 0.041 – RH_925 * 0.239
		<b>CART nodes and multiple regression equations:</b>
	O <sub>3MAX</sub>	<b>For PM<sub>2.5</sub>D12 ≤ 18.85</b> PM <sub>2.5</sub> = PM <sub>2.5</sub> D12 * 0.860 + H_850 * 0.001
		<b>For 18.85 &lt; PM<sub>2.5</sub>D12 ≤ 27.25</b> PM <sub>2.5</sub> = PM <sub>2.5</sub> D12 * 0.975
		<b>For PM<sub>2.5</sub>D12 &gt; 27.25</b> PM <sub>2.5</sub> = PM <sub>2.5</sub> D12 * 0.873
O <sub>3MAX</sub>	<b>For O<sub>3</sub>D12 ≤ 51.20</b> O <sub>3MAX</sub> = V <sub>MAX</sub> * 2.122 – STB_700 * 1.511	
	<b>For 51.20 &lt; O<sub>3</sub>D12 ≤ 75.55</b> O <sub>3MAX</sub> = O <sub>3MAX</sub> D12 * 0.707 + TAIR <sub>MAX</sub> * 2.144 + THI_700 * 1.595	
	<b>For O<sub>3</sub>D12 &gt; 75.55</b> O <sub>3MAX</sub> = O <sub>3</sub> D12 * 0.490 – STB_700 * 1.945 + WW * 9.065	
Madeira Autonomous Region: São João	NO <sub>2MAX</sub>	<b>Multiple regression equation:</b> NO <sub>2MAX</sub> = NO <sub>2</sub> D12 * 0.7950 + H_500 * (0.0030) – DEWP <sub>MEA</sub> * 0.6160 – WW * 1.4930 – Daylight * 0.4520 + TAIR_925 * 0.4380 – V <sub>MAX</sub> * 0.1950 – STB_850 * 0.1370
	PM <sub>10</sub>	<b>Multiple regression equation:</b> PM <sub>10</sub> = PM <sub>10</sub> D12 * 0.9360 + TAIR_925 * (0.4130) – PM10D12 * 0.0960 – DEWP <sub>MEA</sub> * 0.2520 + PM <sub>10</sub> D3 * 0.0650 – WW * 0.9360
	PM <sub>2.5</sub>	<b>CART nodes and multiple regression equations:</b>
		<b>For PM<sub>2.5</sub>D12 ≤ 5.95</b> PM <sub>2.5</sub> = PM <sub>2.5</sub> D12 * 0.895 + TAIR_925 * 0.060
		<b>For 5.95 &lt; PM<sub>2.5</sub>D12 ≤ 9.55</b> PM <sub>2.5</sub> = PM <sub>2.5</sub> D12 * 1.079 – WW * 0.407 – RH_925 * 0.006
		<b>For PM<sub>2.5</sub>D12 &gt; 9.55</b> PM <sub>2.5</sub> = PM <sub>2.5</sub> D12 * 0.814 – WW * 1.703 + TAIR <sub>MAX</sub> * 0.122
Macao Administrative Region: Taipa Ambient	NO <sub>2</sub>	<b>Multiple regression equation:</b> NO <sub>2</sub> = NO <sub>2</sub> _D16 * 0.914 + H_850 * 0.004 + STB_925 * 0.734
	PM <sub>10</sub>	<b>Multiple regression equation:</b> PM <sub>10</sub> = PM <sub>10_16D1</sub> * 0.905 + H_850 * 0.014 – RH <sub>MEA</sub> * 0.205
	PM <sub>2.5</sub>	<b>Multiple regression equation:</b> PM <sub>2.5</sub> = PM <sub>2.5_16D1</sub> * 0.928 + H_850 * 0.006 – RH <sub>MEA</sub> * 0.093
	O <sub>3MAX</sub>	<b>CART nodes and multiple regression equations:</b>
		<b>For O<sub>3</sub> MAX_16D1 ≤ 105.50</b> O <sub>3MAX</sub> = O <sub>3MAX_16D1</sub> * 1.034 – O <sub>3MAX_23D1</sub> * 0.214 + H_850 * 0.019 – RH <sub>MIN</sub> * 0.236
		<b>For 105.50 &lt; O<sub>3MAX_16D1</sub> ≤ 181.87</b> O <sub>3MAX</sub> = O <sub>3MAX_16D1</sub> * 0.994 – O <sub>3MAX_23D1</sub> * 0.433 + H_850 * 0.051 – RH <sub>MIN</sub> * 0.529
		<b>For O<sub>3MAX_16D1</sub> &gt; 181.87</b> O <sub>3MAX</sub> = O <sub>3MAX_16D1</sub> * 1.006 – O <sub>3MAX_23D1</sub> * 0.473 – STB_850 * 8.608



**TABLE 5** | Model performance indicators for validation with 2019 data, by AQMS and pollutant, at Greater Lisbon Area.

Station	Type	Pollutant	Model performance indicators					Model type
			$R^2$	RMSE	MAE	BIAS	RMAE (%)	
Avenida da Liberdade	Urban traffic	PM <sub>10</sub>	0.79	5.1	3.8	1.1	14.6	CART + MR
		NO <sub>2MAX</sub>	0.62	23.7	18.3	2.6	17.9	CART + MR
Entrecampos	Urban traffic	PM <sub>10</sub>	0.76	5.0	3.6	0.9	16.5	CART + MR
		PM <sub>2.5</sub>	0.50	5.4	3.7	0.4	30.2	CART + MR
		NO <sub>2MAX</sub>	0.71	17.6	13.1	1.0	18.5	CART + MR
		O <sub>3MAX</sub>	0.62	11.6	8.8	2.6	12.8	CART + MR
Olivais	Urban background	PM <sub>10</sub>	0.71	5.3	4.1	1.6	21.3	CART + MR
		PM <sub>2.5</sub>	0.52	4.9	3.5	0.9	34.3	CART + MR
		NO <sub>2MAX</sub>	0.69	20.4	13.2	3.6	20.6	MR
		O <sub>3MAX</sub>	0.64	12.0	8.8	0.6	11.4	MR
Mem Martins	Urban background	PM <sub>10</sub>	0.81	3.1	2.3	0.1	12.9	CART + MR
		PM <sub>2.5</sub>	0.76	2.1	1.7	0.4	20.4	CART + MR
		NO <sub>2MAX</sub>	0.76	13.0	8.3	-0.2	27.4	MR
		O <sub>3MAX</sub>	0.66	10.8	8.0	-0.4	9.2	CART + MR

**TABLE 6** | Model performance indicators for validation with 2019 data, by AQMS and pollutant, at Madeira Autonomous Region.

Station	Type	Pollutant	Model performance indicators					Model type
			$R^2$	RMSE	MAE	BIAS	RMAE (%)	
São João	Urban traffic	PM <sub>10</sub>	0.83	4.6	2.6	0.5	13.6	MR
		PM <sub>2.5</sub>	0.85	1.5	1.0	0.2	13.9	CART + MR
		NO <sub>2MAX</sub>	0.82	7.5	6.0	2.2	14.4	MR
São Gonçalo	Urban background	PM <sub>10</sub>	0.70	7.4	3.6	-0.2	23.6	CART + MR
		O <sub>3MAX</sub>	0.67	12.1	9.6	-3.2	9.5	MR

**TABLE 7** | Model performance indicators for validation with 2019 data, by AQMS and pollutant, at Macao Administrative Region.

Station	Type	Pollutant	Model performance indicators					Model type
			$R^2$	RMSE	MAE	BIAS	RMAE (%)	
Macao Roadside	Urban traffic	PM <sub>10</sub>	0.88	8.4	5.6	1.5	11.8	MR
		PM <sub>2.5</sub>	0.87	5.2	3.3	0.2	13.6	MR
		NO <sub>2</sub>	0.89	7.9	5.8	-0.1	9.8	MR
Macao Residential	High density residential	PM <sub>10</sub>	0.89	8.8	5.9	-0.1	10.3	MR
		PM <sub>2.5</sub>	0.87	5.2	3.3	0.8	14.0	MR
		NO <sub>2</sub>	0.86	7.7	5.5	0.0	10.9	MR
		O <sub>3MAX</sub>	0.85	23.2	14.0	0.0	22.3	MR
Taipa Ambient	Urban background	PM <sub>10</sub>	0.88	7.8	5.1	0.8	14.3	MR
		PM <sub>2.5</sub>	0.86	4.8	3.1	0.2	17.7	MR
		NO <sub>2</sub>	0.87	6.1	4.2	1.0	16.3	MR
		O <sub>3MAX</sub>	0.86	23.7	14.7	-1.6	13.9	CART + MR
Taipa Residential	High density residential	PM <sub>10</sub>	0.88	7.9	5.1	0.2	8.7	MR
		PM <sub>2.5</sub>	0.88	5.6	3.5	-0.1	13.1	MR
		NO <sub>2</sub>	0.87	5.6	4.1	0.6	12.8	MR
		O <sub>3MAX</sub>	0.78	20.9	12.7	1.3	19.7	CART + MR

send advisory recommendations in the anticipation of higher pollution episodes to avoid excessive exposure.

The CART analysis defines a path with several nodes, where threshold values on specific variables split into binary ramifications, based on the largest reduction in variations in the target variable in each of the new branch (Choi et al., 2013). CART analysis produces a tree representation, as exemplified in **Figure 3** for PM<sub>10</sub>, according to some parameterizations, as the pretended tree depth. The CART analysis and the MR model development were performed using the IBM software SPSS (Version 25).

Model performance evaluation was accomplished by computing the most common scores: (i) bias (Equation 1), (ii) mean absolute error (MAE) (Equation 2), (iii) root mean square error (RMSE) (Equation 3), (iv) coefficient of determination ( $R^2$ ) (Equation 4), and (v) relative mean absolute error (RMAE) (Equation 5), where  $f$  represents the forecasting value,  $o$  the observed value,  $n$  the forecast/observation pairs,  $\bar{f}$  the forecasting mean value, and  $\bar{o}$  the observed mean value. These statistic measures of agreement were obtained by comparing the forecasted 2019 validation data set to the observed/monitored air pollutant levels on that year.

$$BIAS = \frac{1}{n} \sum_{i=1}^n (f_i - o_i) \quad (1)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - o_i| \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - o_i)^2} \quad (3)$$

$$R^2 = \frac{[\sum_{i=1}^n (f_i - \bar{f}) - (o_i - \bar{o})]^2}{2 [\sum_{i=1}^n (f_i - \bar{f})^2] [\sum_{i=1}^n (o_i - \bar{o})^2]} \quad (4)$$

$$RMAE = \left( \frac{MAE}{\bar{o}} \right) * 100 \quad (5)$$

The obtained statistical models allow to perform a daily forecast, for the next day, of PM<sub>10</sub>, PM<sub>2.5</sub>, NO<sub>2</sub>, and O<sub>3</sub> concentrations, in an operational mode, for the three studied regions. The prediction models run daily, after 16 h for Macao and 13 h for both Greater Lisbon Area and Madeira Autonomous Region, according to the daily schedules in which air quality data is made available.

In the final stage of the operational process, a forecasted air quality index (AQI) is produced for each pollutant for the next day based on the daily air pollutant concentrations, mean or maximum daily values, depending on the pollutant. The final AQI, for each location, corresponds to the worst level of air quality among the forecasted pollutants.

## RESULTS

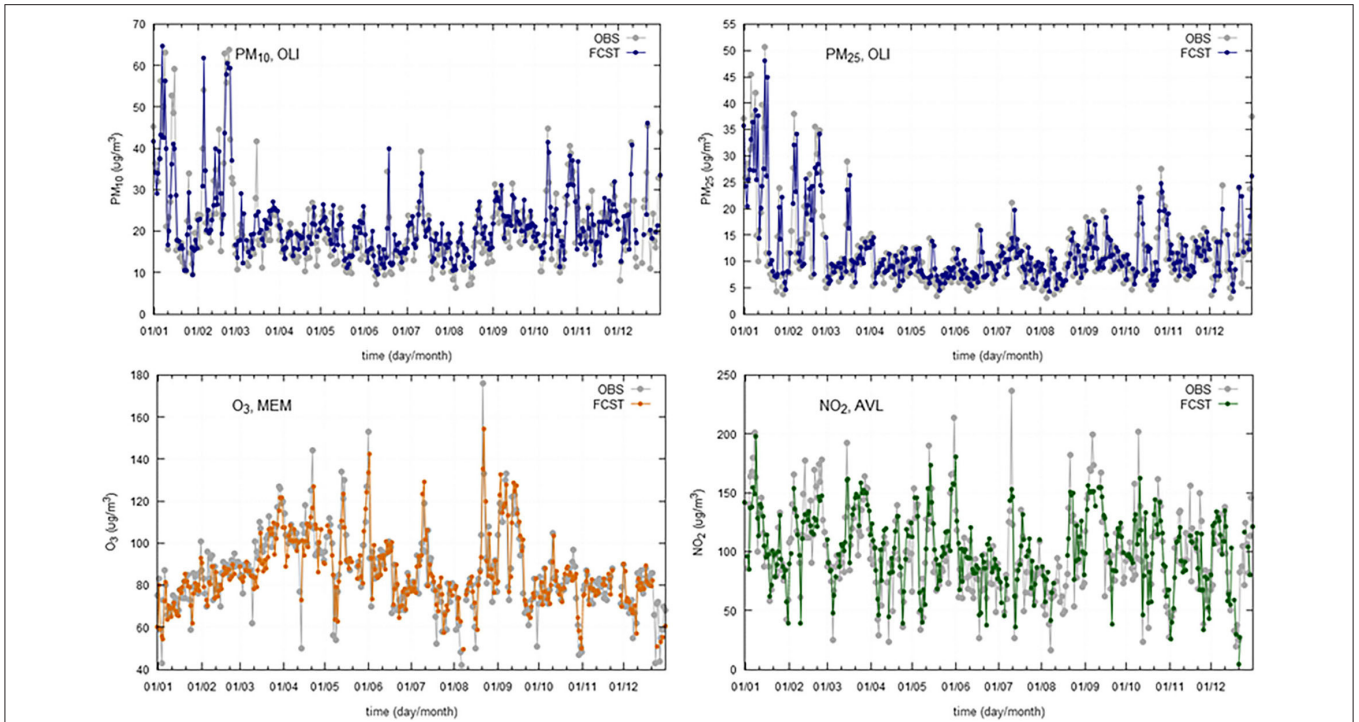
### Model Selected Variables and Performance Indicators

The statistical models based on MR and CART analysis were developed to forecast NO<sub>2</sub>, PM<sub>10</sub>, PM<sub>2.5</sub>, and O<sub>3</sub> concentrations. The objective is to perform a daily forecast, for the next day, in an operational mode by running the prediction models after 16 h for Macao and 13 h both Greater Lisbon Area and Madeira Autonomous Region.

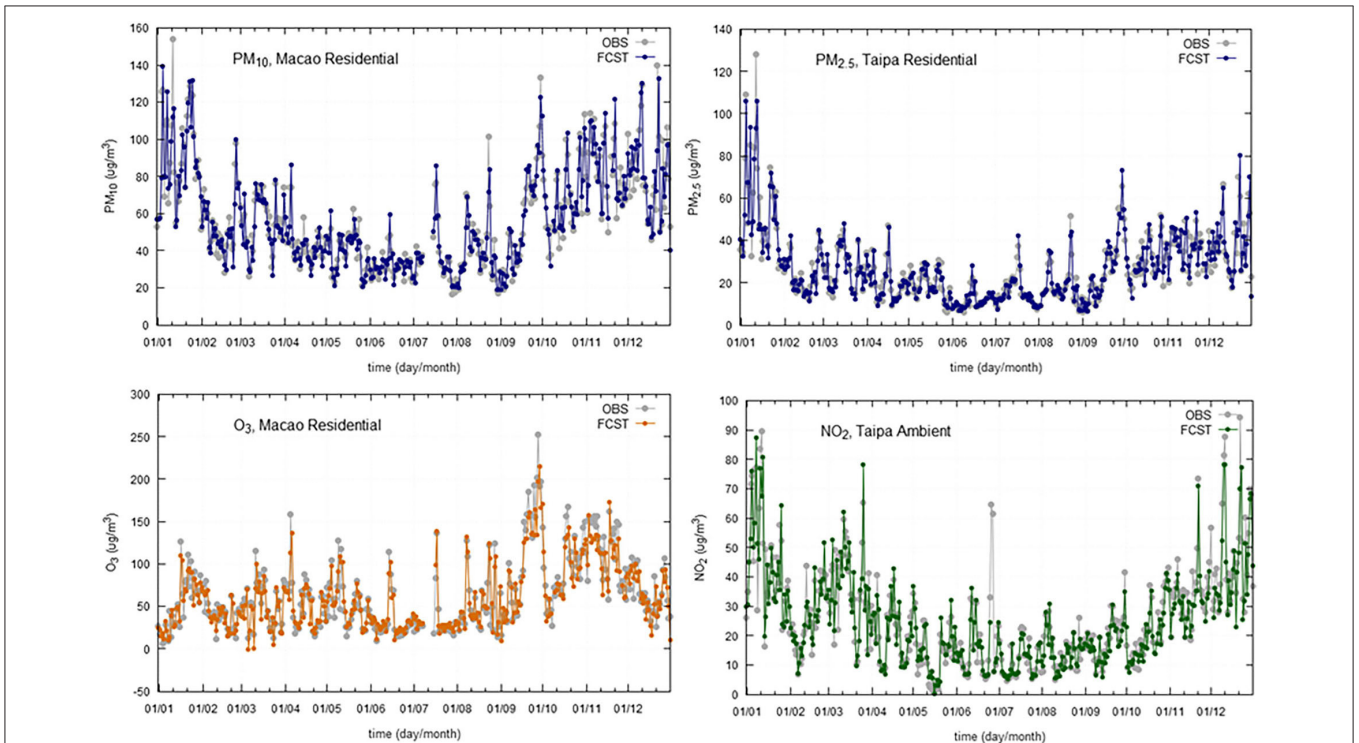
CART analysis was tested mainly to better predict high concentration levels. For Macao region, CART analysis did not improve the quality of overall predictions being, in this case, the prediction models based only on one MR equation. We believe that in Macao, pollution is frequently due to distant sources with pollutants being, transported through the advection of air masses by large scale circulation. Therefore, local meteorology is not as critical, being one equation sufficient to explain and enable the prediction of next day pollutant concentrations, for each monitoring location. The exception was verified for O<sub>3MAX</sub> predictions, at two AQMS. In these last cases, CART analysis allowed to identify split nodes, for which O<sub>3</sub> prediction equations were determined, afterward, by using MR for each node. Opposing to Macao trend, in Greater Lisbon Area, almost every AQMS and pollutant are being forecasted with CART and MR.

The most prevalent variable, being selected at all the forecast equations, is the one that represents the last 24-h pollutant concentrations (16 h from yesterday to 15 h today in the case of Macao, and 11 h of yesterday to 12 h of today for the Great Lisbon Area and Madeira Autonomous Region). Regarding the meteorological selected independent variables used as predictors (**Table 3**), the geopotential height at 850 hPa (H\_850), indicator of synoptic-scale weather pattern, is frequently present in the forecast of NO<sub>2</sub> and PM, both in Lisbon and Macao. For Lisbon and Madeira stations, the most common and frequent weather variable is air temperature at 925 hPa (TAIR\_925), a measure of the strength and height of the subsidence inversion, especially for PM forecasts. In Macao, in addition to H\_850 (the most common variable), figures the RH<sub>MEA</sub>, attributing relevance to relative humidity in the air quality forecast at this region. In Lisbon, the final set of model selected variables covers 13 weather variables. In Madeira and Macao, there is a lower variability of different weather variables: for Madeira TAIR\_925 and the average dew point temperature (DEWP<sub>MEA</sub>) are the most common with more four different variables, and in Macao H\_850 and RH<sub>MEA</sub> are the most common with more seven different variables. In **Table 4**, the obtained MR model equations are presented for one AQMS selected for each studied region.

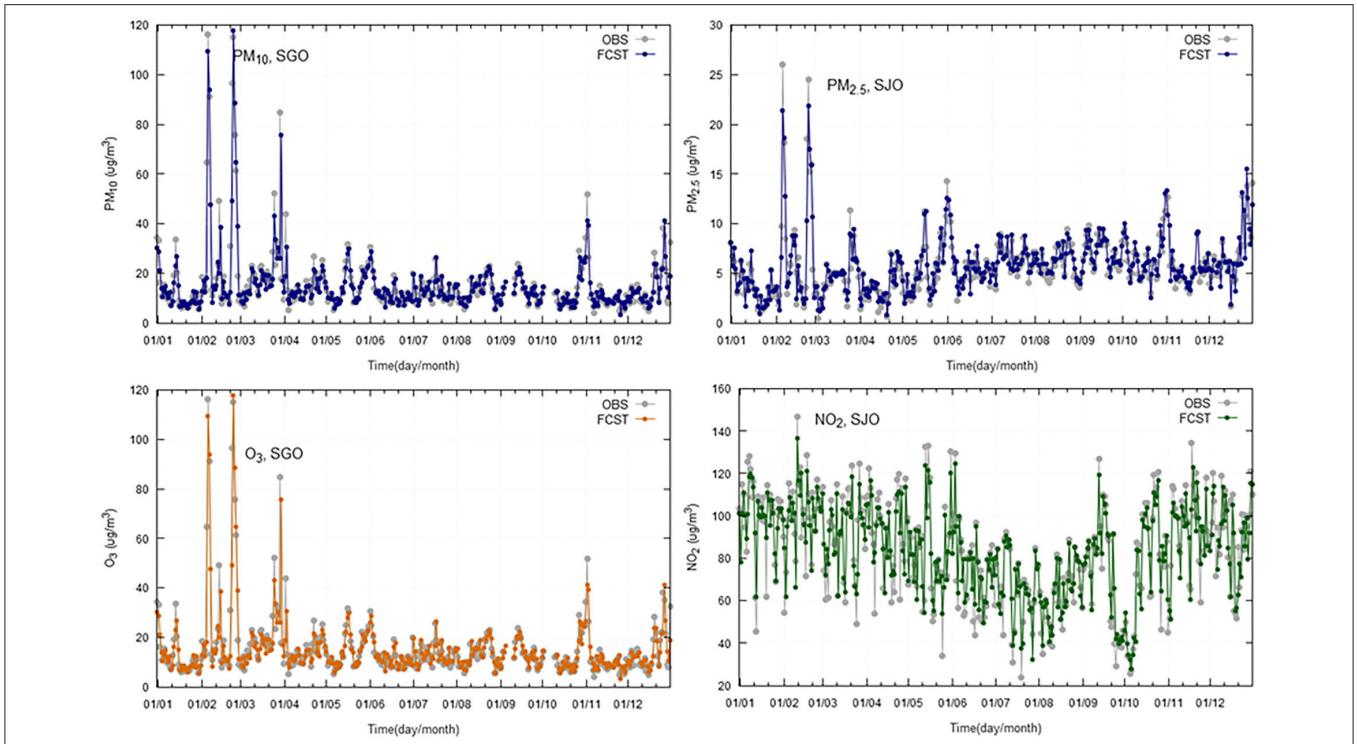
Models were validated with collected data from 2019. For validation purposes, it is important to use at least 1 year of data, to accommodate for all the seasonal variations. Model performance indicators are summarized, by region, in **Table 5** (Greater Lisbon Area), **Table 6** (Madeira Autonomous Region) and **Table 7** (Macao Administrative Region). The referred tables contain the obtained model performance indicators, such as,  $R^2$ , RMSE, MAE, Bias, and RMAE. For each station and pollutant,



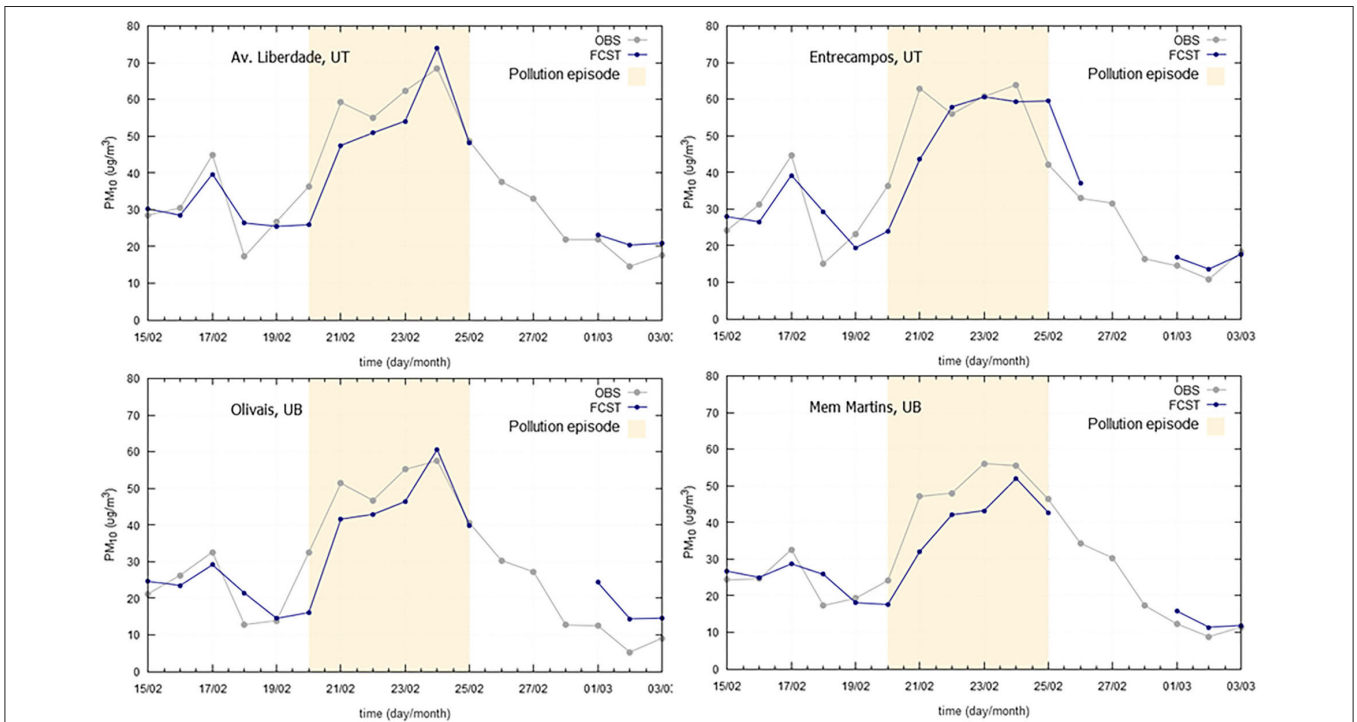
**FIGURE 4 |** Daily observations (OBS) and forecasts (FCST) at three monitoring stations (AVL, Avenida da Liberdade; MEM, Mem Martins; ENT, Entrecampos) in Greater Lisbon Area, for 2019.



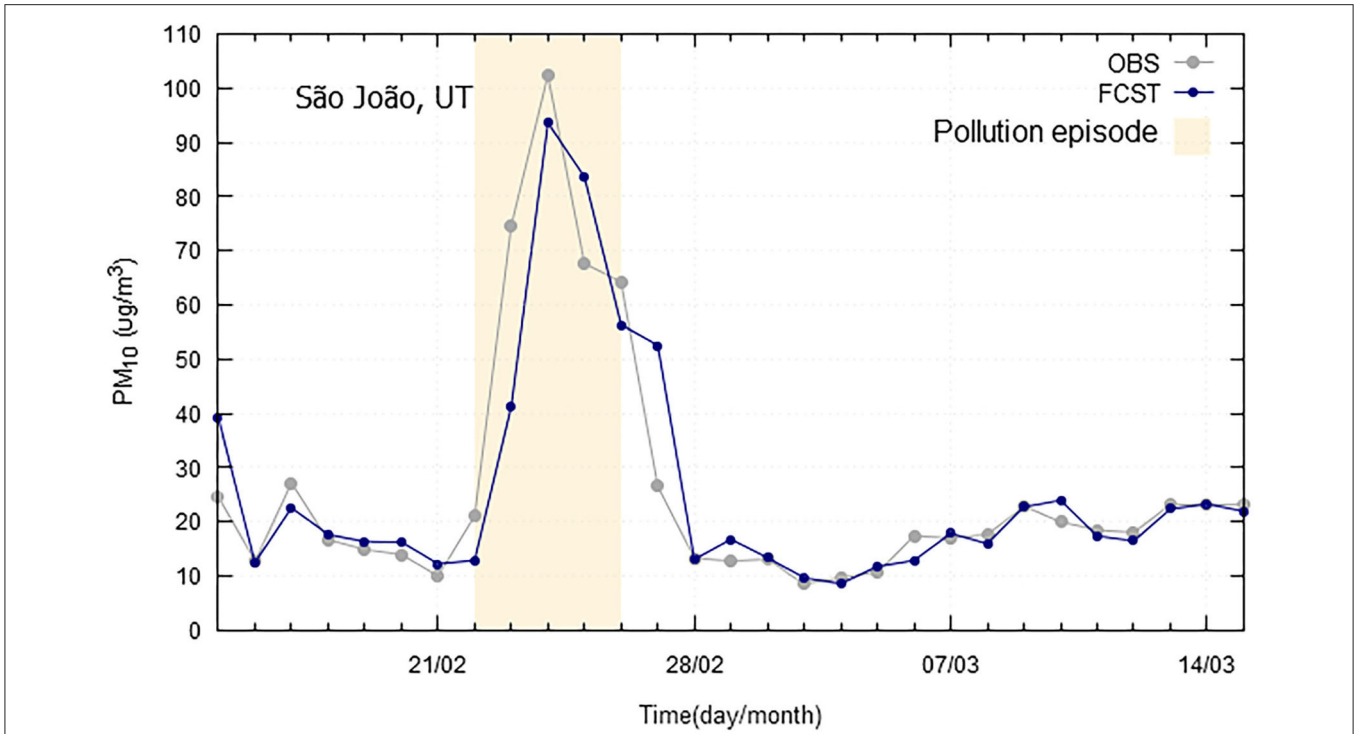
**FIGURE 5 |** Daily observations (OBS) and forecasts (FCST) at three monitoring stations (Macao Residential, Taipa Residential, and Taipa Ambient) in Macao Administrative Region, for 2019.



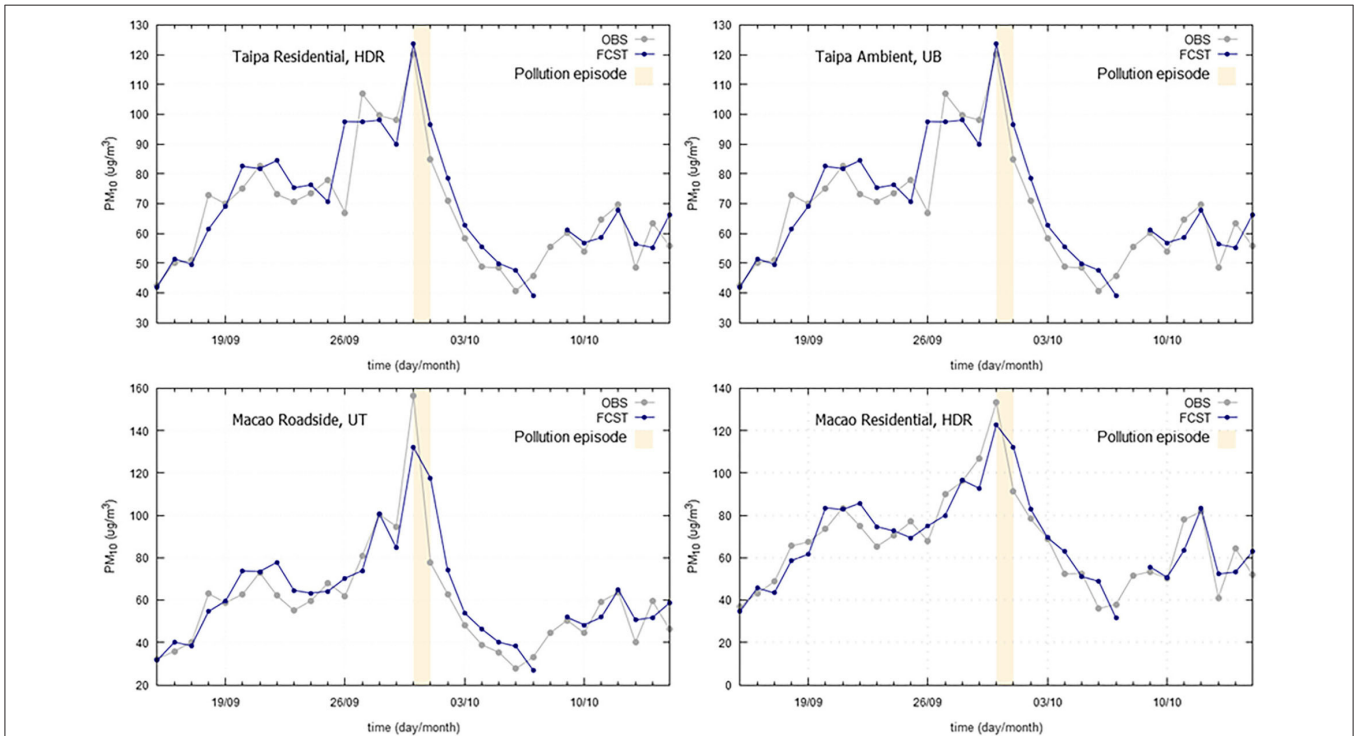
**FIGURE 6 |** Daily observations (OBS) and forecasts (FCST) at two monitoring stations (SJO, São João; SGO, São Gonçalo) in Madeira Autonomous Region, for 2019.



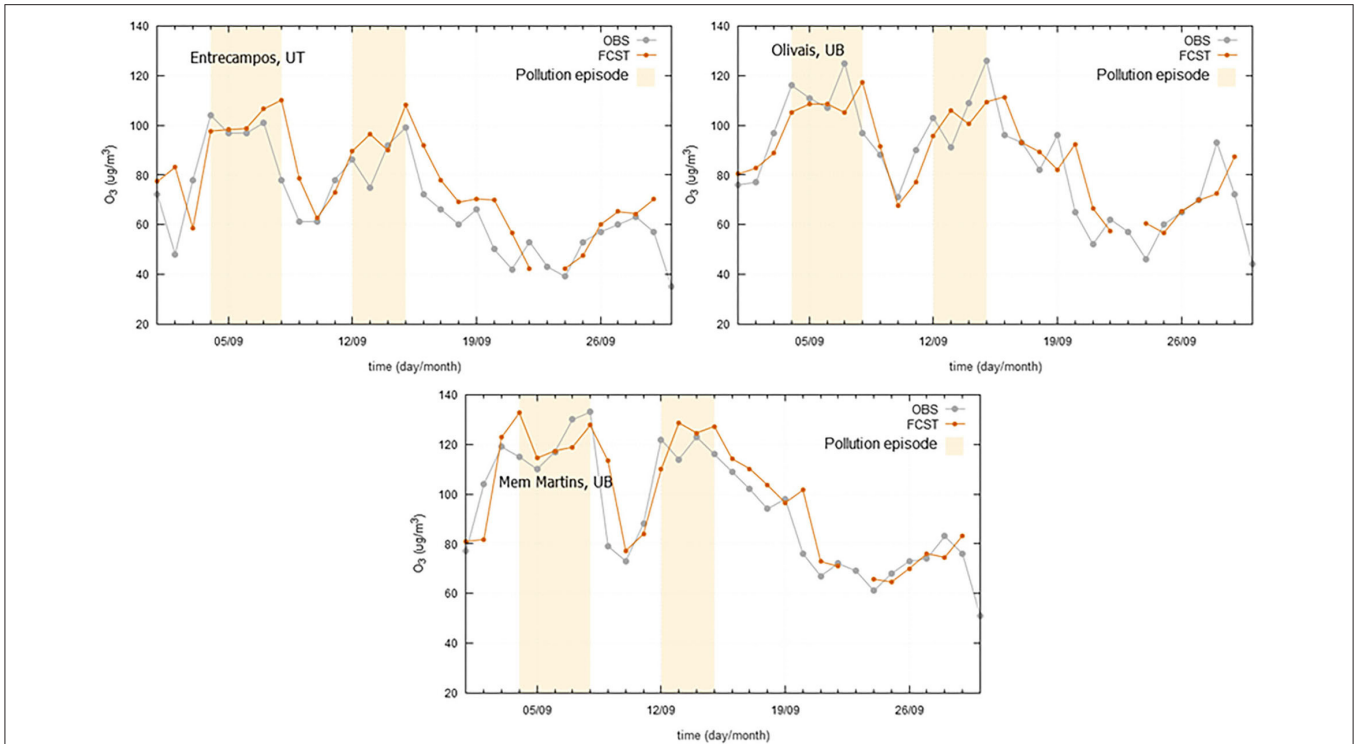
**FIGURE 7 |** Particulate matter (PM<sub>10</sub>) observed (OBS) and forecasted (FCST) concentrations, with emphasis on the natural dust episode occurred in 2019 (20-25/02/2019), at four air quality monitoring stations at Greater Lisbon Area.



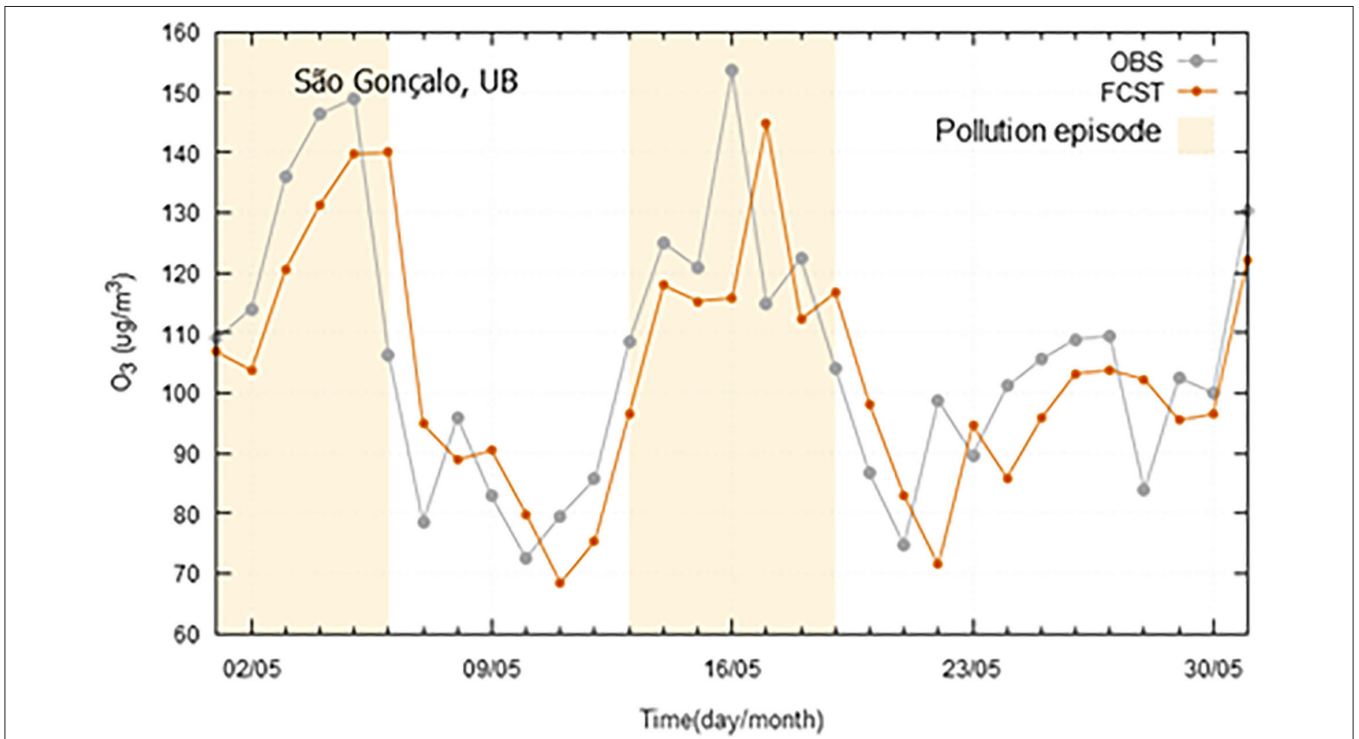
**FIGURE 8** | Particulate matter (PM<sub>10</sub>) observed (OBS) and forecasted (FCST) concentrations, with emphasis on the natural dust episode occurred in 2019 (22-26/02/2019), at São João air quality monitoring station at Madeira Autonomous Region.



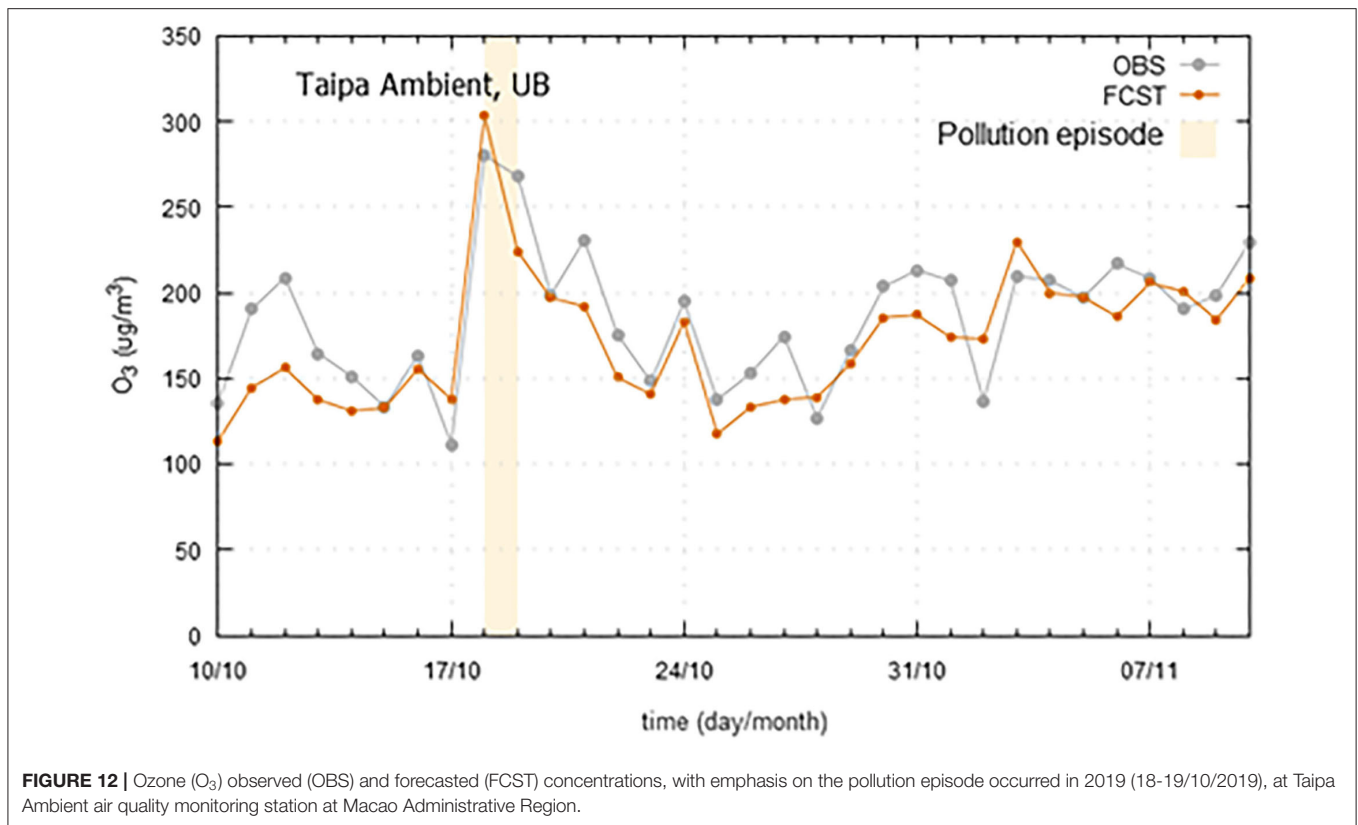
**FIGURE 9** | Particulate matter (PM<sub>10</sub>) observed (OBS) and forecasted (FCST) concentrations, with emphasis on the Chinese National Holiday in 2019 (01/10/2019), at four monitoring stations at Macao Administrative Region.



**FIGURE 10** | Ozone (O<sub>3</sub>) observed (OBS) and forecasted (FCST) concentrations, with emphasis on the pollution episodes occurred in 2019 (04-08/09/2019 and 12-15/09/2019), at three air quality monitoring stations at Greater Lisbon Area.



**FIGURE 11** | Ozone (O<sub>3</sub>) observed (OBS) and forecasted (FCST) concentrations, with emphasis on the pollution episode occurred in 2019 (01-06/05/2019 and 13-19/05/2019), at São Gonçalo air quality monitoring station at Madeira Autonomous Region.



the forecasted time series was plotted against observations (Figures 4–6).

The results show a good agreement between modeled and observed concentrations, being statistically significant at the 95% CI. The selected models provide a good relationship between meteorological and air quality variables, when performing an air quality forecast under different situations.

The time series plotting displays a good overall correlation between observations and forecasted values, however, there is a slight trend to underestimate the maximum peaks. The statistical scores are comparable across the regions under analysis.

Regarding the obtained  $R^2$  for modeled vs. observed concentrations (Tables 5–7), the following aspects can be highlighted:

- the  $R^2$  is, on average, lower for Lisbon than for Madeira and Macao, ranging from 0.5 in Olivais for  $PM_{2.5}$  to 0.81 in Mem Martins for  $PM_{10}$ ;
- Macao presents  $R^2$ -values ranging from 0.78 at Taipa Residential for  $O_3$  to 0.89 at Macao Roadside and Macao Residential (for  $NO_2$  and  $PM_{10}$ , respectively);
- Madeira  $R^2$ -values range from 0.67 at São Gonçalo for  $O_3$  to 0.85 at São João for  $PM_{2.5}$ .

In general, the bias stays very close to zero with the maximum value being 3.6 achieved for the  $NO_2$  at the station Olivais (Lisbon). When comparing the BIAS and MAE, there are significant differences between pollutants, some of them related to different ranges of variation of the daily concentrations. The

RMAE ranges from 8.7% for  $PM_{10}$  at Taipa Residential (Macao) to 34.3% for  $PM_{2.5}$ , at Olivais. Comparing the RMAE for the different regions, Lisbon displays, on average, higher values than Macao and Madeira, being  $PM_{2.5}$  the pollutant with the lowest performance.

## Atmospheric Pollution Episodes

As examples of the response of developed models, in situations where air pollutant concentrations rise significantly, a few pollution episodes were chosen for each region under study, considering different pollutants:  $PM_{10}$  and  $O_3$ .

Long-range transport processes of desert dust from North Africa are not infrequent, significantly affecting ground level particle concentrations recorded during these events, in Iberian Peninsula (Querol et al., 2009). In Portugal, both in Lisbon and Madeira, these natural dust intrusion episodes are common, contributing to higher  $PM_{10}$  concentrations, frequently above the daily limit value of  $50 \mu\text{g}/\text{m}^3$ , as represented in Figures 7, 8, often due to the persistence of specific synoptic patterns. Both in Lisbon and Madeira case studies, forecast models slightly underestimated  $PM_{10}$  concentrations, but were able to follow general  $PM_{10}$  evolution profile, showing a small delay in the prediction trend.

Concerning Macao Administrative Region, a period covering the Chinese National Holiday was chosen, in which a rise of  $PM_{10}$  concentrations, to values over  $120 \mu\text{g}/\text{m}^3$ , was measured at four air quality monitoring stations (Figure 9). This holiday is known to be a golden week of tourism, Macao being one of

the favorite destinations for Chinese tourists (Lee et al., 2017) and also characterized by the release of a considerable amount of fireworks. The PM<sub>10</sub> peak concentrations, occurred on the 1<sup>st</sup> of October, was well predicted for Taipa monitoring locations and slightly underestimated for Macao stations.

Regarding O<sub>3</sub>, a set of pollution episodes, occurred in 2019 in the three case study regions, is presented in **Figures 10–12**. The marked high pollution intervals in these figures correspond to the exceedance of pollutant specific legal thresholds. Mechanisms for near-surface ozone formation and depletion are complex. Previous studies have shown that ozone production accelerates at high temperatures, which may be attributed not only to the temperature dependence of chemical reactions, but also to the weak winds which accompany high temperatures and heatwaves, and cause the atmosphere to stagnate and built up ozone levels (Pyrgou et al., 2018). In a general mode, all the models have shown a good agreement between the observed and forecasted concentrations and were able to forecast the pollution peaks with a good degree of precision. However, in the case of Madeira, due to the particular circumstances of being an island with extreme altitude variations, a meteorological next-day variable was not found to be integrated in the model and anticipate some higher ozone levels. Ozone, as secondary pollutant, has a complex formation process that creates higher forecast difficulties in certain geographical areas. Therefore, in some of these situations, the lag shown in **Figure 11**, between observed and predicted ozone concentrations, is mostly a consequence of the daily evolution trend from the day before.

## DISCUSSION

The described statistical approach to air quality forecasting has proven to be successful, being able to forecast next day NO<sub>2</sub>, PM<sub>10</sub>, PM<sub>2.5</sub>, and O<sub>3</sub> concentrations with a good performance reflected by the presented evaluation scores. The results differ slightly, between stations and pollutant, but overall the variables included in each model explain more than 90% of the variance of the independent variable in the development stage, value that usually decreases in the validation period. The application to different regions pretends to demonstrate the versatility of the methodology. It is expected a small degradation in the performance of the models when in operation, due to several factors, such as the uncertainty of meteorological forecasts.

Statistical models should be updated on a regular basis if there are, for instance, significant changes on local sources of air pollution, but can also be improved with the introduction of new variables, as predictors, in order to better explain part of pollutant variance.

## REFERENCES

AQMD (2022). South Coast AQMD, <https://www.aqmd.gov/home/programs/community/community-detail?title=check-before-you-burn> (accessed November 22, 2021).

The forecast models can show a slight delay in response to the short-term sudden variations on concentrations, once the previous day concentration is itself the independent variable considered as the best predictor since it explains most of the model variance. However, as shown in the selection of PM<sub>10</sub> and O<sub>3</sub> pollution episodes, this methodology was able to reproduce the trend and variations of monitored air pollutant concentrations. This shows that the regression models obtained can be reliably applied to forecast next-day pollutants concentrations across different magnitude levels of air pollution, being a useful tool for air pollution impacts mitigation.

The method has a few advantages when compared with numerical modeling, namely the lower complexity of development and implementation, and the fewer computing resources needed. On the disadvantages side, it can be pointed at the high dependence on a good operating air quality monitoring network and meteorological forecasts. In the case of Portugal, where next day forecasts provided by the Portuguese Environmental Agency are calculated by two methods (deterministic and stochastic) as part of an ensemble approach for both PM<sub>10</sub> and O<sub>3</sub>, the quality data for 2019 show that the probability of detection, by the stochastic model was higher for all the regions (6) within the country except for one.

## DATA AVAILABILITY STATEMENT

The presented datasets belong to public institutions. Requests to access these datasets should be directed to [lc.mendes@fct.unl.pt](mailto:lc.mendes@fct.unl.pt).

## AUTHOR CONTRIBUTIONS

LM, JM, and FF: data curation and writing—review and editing. FF: funding acquisition. LM: methodology, software, and writing—original draft. JM and FF: supervision. LM and JM: validation. All authors have read and agreed to the published version of the manuscript.

## FUNDING

This research is based on the outcomes from the Portuguese PrevQualar Project, supported by the Portuguese Environment Agency. Furthermore, the preparatory work performed for Macao Administrative Region was supported by the Macao's Meteorological and Geophysical Bureau (SMG). The research activities developed at CENSE are financed by the Portuguese Foundation for Science and Technology (FCT) through the Strategic Project UIDB/04085/2020. The presented work was made possible with the support by Portuguese Institute for Sea and Atmosphere (IPMA). Concerning the Macao modelling tasks, the work was supported by Dr. Man Tat-Lei.

Bai, L., Wang, J., Ma, X., and Lu, H. (2018). Air pollution forecasts: an overview. *Int. J. Environ. Res. Public Health* 15, 780. doi: 10.3390/ijerph15040780

Boubel, R. W., Fox, D. L., Bruce Turner, D., and Stern, A. C. (1994). *Fundamentals of Air Pollution, 3rd Edn*. Cambridge, MA: Academic Press.



- Cassmassi, J. C. (1987). Development of an objective ozone forecast model for the South Coast Air Basin. in *Annual Meeting of the Air Pollution Control Association, Conference: 80, Vol. 4* (New York, NY), 21–6.
- Choi, W., Paulson, S. E., Cassmassi, J., and Winer, A. M. (2013). Evaluating meteorological comparability in air quality studies: classification and regression trees for primary pollutants in California's South Coast Air Basin. *Atmos Environ.* 64, 150–159. doi: 10.1016/j.atmosenv.2012.09.049
- Durão, R. M., Mendes, M. T., and Pereira, M. J. (2016). Forecasting O3 levels in industrial area surroundings up to 24 h in advance, combining classification trees and MLP models. *Atmos. Pollut. Res.* 7, 961–970 doi: 10.1016/j.apr.2016.05.008
- Entwistle, M. R., Gharibi, H., Tavallali, P., Cisneros, R., Schweizer, D., Brown, P., et al. (2019). Ozone pollution and asthma emergency department visits in Fresno, CA, USA, during the warm season (June–September) of the years 2005 to 2015: A time-stratified case-crossover analysis. *Air Qual. Atmos. Heal.* 12, 661–672. doi: 10.1007/s11869-019-00685-w
- European Union Legislation (2008). *Directive 2008/50/EC*. Available online at: <https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:32008L0050> (accessed November 20, 2021).
- IPMA (2022). *Portuguese Institute for Sea and Atmosphere*. Available online at: <https://www.ipma.pt/pt/oclima/normais.clima/1971-2000/normalclimate7100.jsp> (accessed November 11, 2021).
- Köppen, W. (1936). Das geographische system der climate. In: *Handbuch der klimatologie*, editors W. Köppen, R. Geiger, Grebrüder Borntraeger, Berlin, 1–44.
- Lee, M., Brauer, M., Wong, P., Tang, R., Tsui, T. H., Choi, C., et al. (2017). Land use regression modelling of air pollution in high density high rise cities: a case study in Hong Kong. *Sci. Total Environ.* 592, 306–315, doi: 10.1016/j.scitotenv.2017.03.094
- Lei, M. T., Monjardino, J., Mendes, L., and Ferreira, F. (2019). Macao air quality forecast using statistical methods. *Air Qual. Atmos. Health* 12, 1049–1057 doi: 10.1007/s11869-019-00721-9
- Lei, M. T., Monjardino, J., Mendes, L., Gonçalves, D., and Ferreira, F. (2020). Statistical forecast of pollution episodes in macao during national holiday and COVID-19. *Int. J. Environ. Res. Public Health* 17, 5124. doi: 10.3390/ijerph17145124
- Liu, H.-Y., Dunea, D., Iordache, S., and Pohoata, A. (2018). A review of airborne particulate matter effects on young children's respiratory symptoms and diseases. *Atmosphere* 9, 150. doi: 10.3390/atmos9040150
- Lopes, D., Hoi, K. I., Mok, K. M., Miranda, A. I., Yuen, K. V., and Borrego, C. (2016). Air quality in the main cities of the pearl river delta region. *Glob. Nest J.* 18, 794–802. doi: 10.30955/gnj.002007
- Neto, J., Torres, P., Ferreira, F., and Boavida, F. (2005). Lisbon air quality forecast using statistical methods. *Int. J. Environ. Pollut.* 39, e028695. doi: 10.1504/IJEP.2009.028695
- NOAA (2001). *National Oceanic and Atmospheric Administration*. Available online at: <https://csl.noaa.gov/aqrsd/reports/forecasting.pdf> (accessed November 22, 2021).
- Oduro, S. D., Ha, Q. P., and Duc, H. (2016). Vehicular emissions prediction with CART-BMARS hybrid models. *Transp. Res. Part D Transp. Environ.* 49, 188–202. doi: 10.1016/j.trd.2016.09.012
- Pyrgou, A., Hadjinicolaou, P., and Santamouris, M. (2018). Enhanced near-surface ozone under heatwave conditions in a Mediterranean island. *Sci. Rep.* 8, 9191. doi: 10.1038/s41598-018-27590-z
- Quarby, S., Santos, G., and Mathias, M. (2019). Air quality strategies and technologies: a rapid review of the international evidence. *Sustainability* 11, 2757. doi: 10.3390/su11102757
- Querol, X., Pey, J., Pandolfi, M., Alastuey, A., Cusack, M., Pérez, N., et al. (2009). African dust contributions to mean ambient PM 10 mass-levels across the Mediterranean Basin. *Atmos. Environ.* 43, 4266–4277. doi: 10.1016/j.atmosenv.2009.06.013
- Seinfeld, J. H., and Pandis, S. N. (2006). *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change, 2nd Edn*. New York, NY: John Wiley and Son, Inc.
- Sheng, N., and Tang, U. W. (2013). Risk assessment of traffic-related air pollution in a world heritage city. *Int. J. Environ. Sci. Technol.* 10, 11–18. doi: 10.1007/s13762-012-0030-1
- SMG (2022). *Climate in Macao*. Available online at: <https://www.smg.gov.mo/en/subpage/124/page/28> (accessed November 10, 2021).
- Tong, C. H. M., Yim, S. H. L., Rothenberg, D., Wang, C., Lin, C.-Y., Chen, Y. D., et al. (2018). Projecting the impacts of atmospheric conditions under climate change on air quality over the Pearl River Delta region. *Atmos. Environ.* 193, 79–87. doi: 10.1016/j.atmosenv.2018.08.053
- US EPA. (2003). *Guidelines for Developing an Air Quality (Ozone and PM2.5) Forecasting Program*. New York, NY: US EPA.
- WHO. (2006). *Air Quality Guidelines: Global Update 2005: Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide*. Geneva: WHO.
- Wiśniewska, K., Lewandowska, A. U., and Staniszevska, M. (2019). Air quality at two stations (Gdynia and Rumia) located in the region of Gulf of Gdansk during periods of intensive smog in Poland. *Air Qual. Atmos. Heal.* 12, 879–890. doi: 10.1007/s11869-019-00708-6
- World Health Organization (2016). *Ambient (Outdoor) Air Pollution Database, By Country and City* (Data file). Available online at: [http://www.who.int/phe/health\\_topics/outdoorair/databases/cities/en/](http://www.who.int/phe/health_topics/outdoorair/databases/cities/en/) (accessed November 20, 2021).
- World Health Organization (2021). *WHO Global Air Quality Guidelines. Particulate Matter (PM2.5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide*. Geneva: World Health Organization.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Mendes, Monjardino and Ferreira. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.