Modeling and forecasting of monthly PM_{2.5} emission of Paris by periodogram-based time series methodology

Yılmaz Akdi · Elif Gölveren ·
Kamil Demirberk Ünlü · Mustafa Eray Yücel

Received: 2 July 2021 / Accepted: 17 August 2021 / Published online: 3 September 2021
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

Abstract In this study, monthly particulate matter (PM_{2.5}) of Paris for the period between January 2000 and December 2019 is investigated by utilizing a periodogram-based time series methodology. The main contribution of the study is modeling the PM_{2.5} of Paris by extracting the information purely from the examined time series data, where proposed model implicitly captures the effects of other factors, as all their periodic and seasonal effects reside in the air pollution data. Periodicity can be defined as the patterns embedded in the data other than seasonality, and it is crucial to understand the underlying periodic dynamics of air pollutants to better fight

pollution. The method we use successfully captures and accounts for the periodicities, which could otherwise be mixed with seasonality under an alternative methodology. Upon the unit root test based on periodograms, it is revealed that the investigated data has periodicities of 1 year and 20 years, so harmonic regression is utilized as an alternative to Box-Jenkins methodology. As the harmonic regression displayed a better performance both in and out-of-sample forecasts, it can be considered as a powerful alternative to model and forecast time series with a periodic structure.

Keywords Harmonic regression · Air pollution · Time series analysis · Periodicity

Y. Akdi
Department of Statistics, Faculty of Science, Ankara
University, Ankara, Turkey
e-mail: akdi@ankara.edu.tr

E. Gölveren
Department of Econometrics, Faculty of Economics
and Administrative Sciences, Ataturk University, Erzurum,
Turkey
e-mail: egolveren@atauni.edu.tr

K. D. Ünlü (✉)
Department of Mathematics, Atilim University, Ankara,
Turkey
e-mail: demirberk.unlu@atilim.edu.tr

M. E. Yücel
Department of Economics, Faculty of Economics,
Administrative and Social Sciences, İhsan Dogramaci
Bilkent University, Ankara, Turkey
e-mail: eray@bilkent.edu.tr

Introduction

Air pollution is one of the most serious environment-related issues that harm ecosystems and human health. Many studies have been conducted to reveal the negative effects of air pollution on human health, and it has been found that exposure to pollutants causes exacerbation of diseases and premature deaths. Evidences suggest that particulate matter is a significant pollutant that mostly increased the risk of cardiovascular and respiratory diseases and lung cancer. In the World Health Organization (WHO) European Region alone, exposure to particulate matter reduces the life expectancy of every person by

an average of approximately 1 year (WHO, 2020b). Therefore, the issue of air pollution and particulate matter pollution remains important.

There are many factors threatening clean air in urban areas where population density is high and modern industrialization is widespread. Therefore, in these regions, it is important to accurately and systematically forecast the air quality and to take the necessary precautions in advance to keep the pollutants below the specified standards. In this study, particulate matter pollution in the Ile de France region, which includes Paris, is evaluated.

PM (particulate matter or particle pollution) is a mixture of solid particles and liquid droplets suspended in the air. Such as dust, dirt, soot, smoke, sea salt, pollen, and spore, it includes combustion by-products and natural particles. It is known that particulate matter has negative effects on allergic inflammation, the lung, the immune system, the heart, the vascular function, and the brain (Breyse et al., 2013; Saxon & Diaz-Sanchez, 2000). Besides, particularly 2.5 μm and less fine particulate matter (PM_{2.5}) air pollution ranks among the leading causes of worldwide morbidity and mortality (Brook et al., 2017). World Health Organization attributes 4.2 million deaths every year to ambient air pollution (WHO, 2020a). European Society of Cardiology and also the American Heart Association have formally recognized ambient PM_{2.5} pollutions as a major cardiovascular risk factor (Brook et al., 2017).

Cases of extreme air pollution observed at various times since the 1930s have created a necessity to examine the effects of air pollution on health. Since the 1970s, more quantitative research has been started on the subject, and more meticulous research has been conducted over time in conjunction with the developments in the techniques used in statistical analyses (Dockery & Pope, 1994).

In the literature, there are various epidemiological studies that investigated the effects of air pollution on mortality and morbidity. These studies have reported that mortality/morbidity rates correlate with particulate air pollution (For related review articles, see: Dockery & Pope, 1994; Ostro et al., 1998; Brunekreef & Holgate, 2002; Pope & Dockery, 2006; Dockery, 2009; Peled, 2011; Stanek et al., 2011; Anderson et al., 2012; Breyse et al., 2013; Hoek et al., 2013; Lippmann, 2014; Schikowski et al.,

2014; Pui et al., 2014; Kim et al., 2015; Mannucci & Franchini, 2017; Liu et al., 2018; Shou et al., 2019).

In the most of studies, it has been reported that there is a positive association between air pollution (particulate matter) and wheezing and allergic rhinitis (Weiland et al., 1994), asthma (Evans et al., 2014; Gasana et al., 2012; Lin et al., 2002; Loftus et al., 2015), lung diseases (Hamra et al., 2014; Maji et al., 2017; Rojas-Bracho et al., 2000), mortality (Schwartz et al., 1996), and early deaths (Apte et al., 2015; Lelieveld et al., 2015; Maji et al., 2018; Orru et al., 2011).

In a study conducted by Dockery et al. (1993), 8111 adults from six cities in the USA were followed up during 14 to 16 years. It is found that air pollution based on fine particulate matter, or an air pollution mixture associated with fine particulate matter, contributes to excess mortality in US cities which is considered in the study. Also, a positive relationship between death from lung cancer and cardiopulmonary disease and air pollution is revealed. Pope et al. (2002) find that long-term exposure to combustion-related fine particulate air pollution is a significant threat to cardiopulmonary and lung cancer mortality, based on data collected from a study which conducted by the American Cancer Society and enrolled approximately 1.2 million adults. Guaita et al. (2011) find an association between daily respiratory-cause mortality and PM_{2.5} concentrations in their study covering the 2003–2005 period in the city of Madrid. Turner et al. (2011) conduct a study based on the CPS-II which is a prospective study. The CPS-II is a study of nearly 1.2 million participants who covered over 77,000 volunteers in 1982. In the study, vital status has been followed up every 2 years. The authors reported that lung cancer mortality is positively associated with mean long-term ambient fine particulate matter air pollution concentrations in the population of lifelong never-smokers. Hart et al. (2015) follow 108,767 members of the Nurses' Health Study (NHS), which is a long-term prospective cohort study of US female nurses, during the 2000–2006 years. They find an association between increasing exposure to PM_{2.5} and an increased risk of mortality. Han et al. (2017) examine lung cancer morbidity from PM_{2.5} concentration exposure in Chinese males for the 2010 to 2015

period. The authors find a significant relationship between lung cancer morbidity and $PM_{2.5}$ concentration level in this study group.

Investigation of the origin of $PM_{2.5}$ has an important place in the literature. For example, Hadley (2017) finds out the primary source factors of $PM_{2.5}$ for the Northwestern United States. The results indicate that the primary sources are ships, wood burning, dust, and sea salt. Also, Rao et al. (2018) show that wood combustion in Italy contributes to the concentration of mean $PM_{2.5}$ by almost 20%. Golly et al. (2019) investigate the organic makers of $PM_{2.5}$ of 5 rural sites of France. They show that wood burning contributes approximately 4 to 10%. Monthly, daily, and hourly $PM_{2.5}$ data of Houston, USA, is studied by Du et al. (2019) from 2008 to 2017. Their findings are summarized as follows: (1) Annual average data has a decreasing trend. (2) Summer months have a higher concentration than the winter months. (3) For the downtown concentration is higher on weekdays while lower for weekends. For rural, suburban, residential, and coastal area the relation is vice versa. (4) Traffic rush hour has a significant effect on concentration. Querol et al. (2001) study the daily concentration level of Barcelona from June 1999 to June 2000. The results show that daily cycles in the level of concentration are consistent with the daily cycle of gaseous pollutants emitted by traffic. Shahid et al. (2016) study $PM_{2.5}$ levels in Karachi, Pakistan. It is shown that calcareous dust and siliceous dust are the dominant material in the PM levels. Also, biomass burning has a share of 3%, while bioaerosol has a minor contribution of 1% to $PM_{2.5}$ levels.

Also, there exists an intensive literature on modeling and forecasting $PM_{2.5}$ emissions. Here, we give some recent studies on air pollution. Zhou et al. (2019) propose a novel multi-output support vector machine (MM-SVM) and multi-task learning algorithm to model and forecast the $PM_{2.5}$ concentrations of Taipei City of Taiwan between 2010 and 2016. They compare the proposed model MM-SVM with a benchmark model, single-output support vector machine (S-SVM), and the results show that MM-SVM performs more prominently than S-SVM. Kow et al. (2020) employ convolutional and backpropagation neural networks (CNN-BP) to forecast the hourly $PM_{2.5}$ concentration of Taiwan during

2017. The results show that CNN-BP is better than a backpropagation neural network. Another study on Taiwan is conducted by Chang et al. (2020a). The researchers propose a backpropagation neural network (BPNN) to model various time scales of $PM_{2.5}$. They conclude that machine learning algorithms are useful to reveal spatiotemporal features of the data and in forecasting. Also, Chang et al. (2020b) propose various machine learning algorithms to model and forecast hourly $PM_{2.5}$ of Taiwan's data. They prove the superiority of the hybrid models.

The hourly $PM_{2.5}$ concentration of Beijing, China, is investigated by Bai et al. (2019a). The authors propose an ensemble long short-term memory neural network (E-LSTM) to forecast the time series data which consists of three steps. The results show that E-LSTM outperforms both single LSTM and feed-forward neural networks in terms of mean absolute percentage error. Hourly data of a Northern China city, Taiyuan, is investigated by Wang et al. (2019) for the periods between 2016 and 2017. The authors utilize a hybrid wavelet model. Their results imply that the forecasting accuracy of artificial neural networks (ANN) and support vector machine (SVM) on the decomposed data has better forecasting accuracy compared to raw data. Another wavelet transformation is proposed by Cheng et al. (2019) to forecast daily $PM_{2.5}$ concentrations of five different cities of China. The transformed data is modeled by autoregressive integrated moving average (ARIMA), SVM, and ANN. The results coincide with the findings of Wang et al. (2019); that is, the models on the transformed data have higher accuracy. Daily $PM_{2.5}$ and O_3 levels of three big cities of China, Beijing, Nanjing, and Guangzhou, are investigated by Lv et al. (2016) by employing a non-linear regression model, while hindcast mode is used as a benchmark model. Their model on forecasting the 2015 values is good in Beijing and Nanjing for both $PM_{2.5}$ and O_3 , but in Guangzhou, only the forecast performance of $PM_{2.5}$ is better than the benchmark model. Yuan et al. (2019) employ novel multi-factor and multi-scale methods to forecast daily $PM_{2.5}$ concentrations of Cangzhou, China, for the periods 2017 and 2018. Their results indicate that the proposed model outperforms popular and similar benchmarks. Classification and regression tree (CART) and ensemble extreme learning

Time Series of PM_{2.5}

Autocorrelation function of PM_{2.5}

Partial Autocorrelation function of PM_{2.5}

◀ **Fig. 1** Time series graph of the $PM_{2.5}$

machine (EELM) methods are proposed by Shang et al. (2019) to model hourly $PM_{2.5}$ concentrations of Yancheng, China, for the periods of 2015 and 2018. The experimental results reveal that the proposed model which addresses global–local duality outperforms the global models, namely, random forest (RF), v-support vector regression (v-SVR), EELM, and other local models on season and k-means clustering. Bai et al. (2019b) propose a seasonal stacked autoencoder model which also takes account on seasonal analysis with deep feature learning to forecast the hourly $PM_{2.5}$ concentration of Beijing for the periods between January 2016 and December 2016. The proposed model is compared with a shallow framework with seasonality consideration (SL-SFN), a deep framework without seasonality consideration (DL-SAE), and shallow framework without seasonality consideration (SL_FFNN). The findings show that the proposed model outperforms the benchmark models. Spatiotemporal variation of hourly $PM_{2.5}$ concentration of Jing-Jin-Ji, China, for the periods between January 2015 and April 2016 is investigated by Qi et al. (2019). The authors propose a hybrid model which consists of graph convolutional networks and long short-term memory networks (GC-LSTM). Several models are used to compare the performance of the model, and it is showed that GC-LSTM outperforms the benchmark models. Hourly $PM_{2.5}$ concentration of Beijing-Tianjin-Hebei area, China, between 2000 and 2019 is studied by Xu et al. (2020). The authors propose a hybrid machine learning model which consists of two parts: a site prediction model and a grid prediction model. The model has better prediction performance compared to benchmark models for the next 24 hours forecasting results. Zhang et al. (2018) study the trends in $PM_{2.5}$ concentration of Fuzhou, China, by employing ARIMA for the periods between August 2014 and July 2016. The ARIMA results show that $PM_{2.5}$ concentrations have seasonal fluctuations of 2 years; higher concentration corresponds to cold periods, while lower concentration corresponds to warm periods.

Perez and Gramsch (2016) investigate the hourly $PM_{2.5}$ concentration of Santiago, Chile, by using multi-layer neural networks for the periods between 2010

and 2012. Their model gives accurate results even 1 hour in advance. Also, Moisan et al. (2018) propose dynamic multi-linear equations (DME) to predict the hourly $PM_{2.5}$ concentration of Santiago. The study also uses Fourier coefficients to capture both annual cycles and the day-of-the-week effect. Another study on Chile is done by Perez et al. (2020) in which the power of ANN is compared with the linear models. $PM_{2.5}$ of Bogota, Colombia is studied by Franceschi et al. (2018) by employing data mining algorithms of ANN, PCA and k-means clustering. It is shown that clustering increases the performance of the algorithms. Another study of $PM_{2.5}$ of Colombia is done by Lopez-Restrepo et al. (2020). The authors use data assimilation methods which are based on ensemble Kalman filter with covariance localization. The investigated periods consist of March–April 2016. The results reveal that data assimilation reduces uncertainty in $PM_{2.5}$ and increases the accuracy. Gibergans Bâguena et al. (2020) propose composition data analysis (CDA) to increase the pollutants knowledge of Barcelona, Spain. The results showed that more powerful air quality index can be formed by using CDA.

Biancofiore et al. (2017) employ recursive artificial neural network, a feed-forward neural network, and a multiple linear regression to model the daily average $PM_{2.5}$ concentration of Pescara, Italy. The comparison of the models shows that the recursive neural network model has better performance.

Akdi et al. (2020a) study monthly PM_{10} of Ankara, the capital city of Turkey, for the periods 1993 and 2017 by employing harmonic regression model. Results of this study indicate that the investigated time series data has periods of 6 months and a year. Also, it is shown that harmonic regression models with trigonometric components outperform the classical ARIMA type model. Okkaoğlu et al. (2020) use periodogram-based time series analysis to enlighten the hidden periodicities of the daily PM_{10} data of London between the periods of 2014 and 2018. Their results indicate that London's data has 7-day, 25-day, 6-month, 1-year, and 15-month periodicities. These periodic structures are used to estimate harmonic regression model, and results show that harmonic regression outperforms other autoregressive type time series models. In above-mentioned studies, although the same methodology

Table 1 ANOVA test result

Source	Degree of freedom	Sum of squares	Mean square	F value	Pr > F
Model	11	2056.171798	186.924709	7.99	< .0001
Error	228	5336.181905	23.404307		
Corrected Total	239	7392.353702			
Months	N		Mean	Standard deviation	
1	20		18.0399450	6.06985452	
2	20		19.0793400	5.35141571	
3	20		200.0374400	7.66584228	
4	20		16.1053350	6.06976783	
5	20		12.7737850	2.76004435	
6	20		12.4429550	2.74112622	
7	20		11.2920450	2.93786745	
8	20		10.7823100	3.18906006	
9	20		12.9993300	3.16359080	
10	20		15.0134050	4.44635087	
11	20		15.2023350	4.00632884	
12	20		16.8112350	6.05289461	

is used, we have fitted different harmonic regressions and found different periodic structure. Each region differs due to its socio-cultural and geographical structure. Each location has its own dynamics; although the proposed model is valid worldwide, local investigations should be carried out to understand local dynamics. By using the local dynamics, the fight against the air pollutants can be carried out from local to general. For that reason, the goal of this study is to model and investigate the monthly $PM_{2.5}$ of Paris for the periods between January 2000 and December 2019 by utilizing periodogram-based time series methodology. Periodogram-based time series methodology is chosen to detect the hidden periodic structure of the data. To the best of our knowledge, this is the first study which investigates the periodic structure of monthly $PM_{2.5}$ of Paris. It is important to reveal these periodic structures. These are hard to find and mostly ignored by researchers and policy

makers. These carry out very important information of the underlying structure of the air pollutants. This information can be used as the measurement of effectiveness of the measures taken. Also, it should be noted that this information is solely extracted from the data itself which increase the efficiency and decrease the complexity of the model. It is shown that the data contains 1-year and 20-year periodicity. Harmonic regression is utilized to model and to forecast data because of the periodic structure. It is shown that harmonic regression has better performance compared to classical time series models.

Contribution to the recent literature

Against the background provided earlier, contribution of this study to the literature can be enumerated as follows:

Table 2 AIC values of the considered models with parameter estimation of σ^2

P	(1)	(2)	(3)	(4)	(6)	(12)	(13)
AIC	1441.946	1443.047	1444.491	1445.987	1437.702	1408.07	1409.457
$\hat{\sigma}_n^2$	23.61623	23.62716	23.67235	23.72321	22.73298	19.61778	19.65448
P	(4)	(12)	(1,4,12)	(1,5,12)	(1,6,12)	(1,4,8,12)	(1,4,6)
AIC	1507.595	1447.778	1409.833	1403.989	1410.042	1410.437	1445.298
$\hat{\sigma}_n^2$	31.04606	24.16691	20.48925	19.99638	20.50708	20.45684	23.75208

Table 3 ADF test result

	<i>t</i> -statistics	Probability
Augmented Dickey-Fuller test statistics	− 9.037172	0.0000
Test critical values:	− 3.457630	
	− 2.873440	
	− 2.573187	

- i. The monthly PM_{2.5} of Paris is predicted and modeled by utilizing periodogram-based time series methodology.
- ii. The proposed model, harmonic regression, only uses the information obtained purely from the time series. Thus, it is efficient, simple, and powerful.
- iii. The hidden periodic structure of monthly PM_{2.5} of Paris is revealed as a year and 20 years.
- iv. The harmonic regression is compared with an ARIMA type methodology, and it is shown that harmonic regression is superior. Thus, it can be considered as a powerful alternative when a time series contains periodic structure.

The rest of the paper is as follows. In Sect. 2, we introduce the methodology. Section 3 is devoted to the data and analysis. Lastly, Sect. 4 concludes this work.

Materials and methods

In this study, we use the methodology proposed by Akdi et al. (2020a) which use periodogram-based time series to model monthly air pollution of Ankara, the capital city of Turkey. The recent literature shows that periodogram-based time series has many applications in economics Akdi et al. (2020b); in weather pollutions Okkaoğlu et al. (2020); in energy markets Akdi et al. (2020c); and in meteorology Akdi and Ünlü (2020). The methodology has many advantages, but the most significant one can be shown that it does not require any assumption or parameter estimations except for the variance of the white noise series. In this section, we summarize the methodology; more details can be found in Akdi et al. (2020a).

Table 4 Parameter estimation of ARIMA model given in Eq. (10)

Parameter	Estimate	Standard error	<i>t</i> -value	Pr> t
α_0	14.94132	0.65818	22.70	< 0.001
α_1	0.38383	0.05544	6.92	< 0.001
α_5	− 0.13220	0.05369	− 2.46	0.0145
α_{12}	0.32964	0.05759	5.72	< 0.001

In general, periodograms are used to identify hidden periodicities found in the time series Fuller (1996), Wei (2006), and Brockwell and Davis (1987). Akdi and Dickey (1998) proposed a test based on periodograms to investigate the unit root which is given as

$$T_n(w_k) = \frac{2(1 - \cos(w_k))}{\hat{\sigma}_n^2} I_n(w_k). \quad (1)$$

Here, the asymptotic behavior of the test statistic is $T_n(w_k) \xrightarrow{D} \chi_1^2 + 3\chi_1^2$ as $n \rightarrow \infty$ where Z_1^2 and Z_2^2 are random variables, which has standard normal distribution. Also, it should be noted that w_k represents the frequency when $w_k = 2\pi k/n$, and then, the Fourier frequencies are obtained. In short, the distribution of the test statistics is (Fuller, 1996)

$$T_n(w_k) \xrightarrow{D} \chi_1^2 + 3\chi_1^2, n \rightarrow \infty. \quad (2)$$

Although the asymptotic distribution is valid for all k values, we set k equals to 1. Let us denote the test statistics as $t_n(w_k)$, and if $t_n(w_k)$ is less than the critical values (c_α), we say that the null hypothesis of time series has unit root is rejected. Critical values of the test are given in Akdi and Dickey (1998). Here, we list critical values of 1%, 5%, and 10% as

Table 5 Forecasting results of ARIMA

Period	Forecast	Observation
2020:01	14.7155	14.0541
2020:02	17.9692	7.6619
2020:03	16.5284	12.0619
2020:04	16.1401	13.2102
2020:05	14.3352	8.9488
2020:06	12.6685	7.4387
2020:07	11.6467	6.4281

Fig. 2 Predictions of ARIMA (black) and observations (blue)

$$\begin{aligned} P(T_n(w_1) \leq 0.0348) &= 0.01, P(T_n(w_1) \\ &\leq 0.178) = 0.05, P(T_n(w_1) \\ &\leq 0.368) = 0.10. \end{aligned} \quad (3)$$

If $t_n(w_k) < c_\alpha$, we say that the investigated time series is stationary under $\alpha\%$ significance level. Periodograms are calculated through trigonometric transformations. In order to test if a time series consists of periodic component or not, for any time series $\{Y_1, Y_2, \dots, Y_n\}$ define the following trigonometric model

amplitude, frequency, and phase, respectively. By taking $w_k = 2\pi k/n$ and using the properties of trigonometric functions, we can rewrite trigonometric model given in Eq. (4) as

$$Y_t = \mu + \alpha \cos(w_k t) + \beta \sin(w_k t) + e_t, t = 1, 2, \dots, n. \quad (5)$$

Here, we use the characteristic of cosine and sinus functions of $\alpha = R \cos(\phi)$ and $\beta = R \sin(\phi)$. Least squares estimators of parameters of the model given in Eq. (5) are

$$\hat{\mu} = \bar{Y}_n, a_k = \frac{2}{n} \sum_{t=1}^n (Y_t - \bar{Y}_n) \cos(w_k t) \text{ and } b_k = \frac{2}{n} \sum_{t=1}^n (Y_t - \bar{Y}_n) \sin(w_k t), \quad (6)$$

$$Y_t = \mu + R \cos(w_k t + \phi) + e_t, t = 1, 2, \dots, n, \quad (4)$$

where e_t is an independent and identically distributed random variable with mean 0 and variance σ^2 , and μ, R, w_k and ϕ are referred to as expected value,

where a_k and b_k are called as Fourier coefficients.

According to the model given in Eq. (5), if the null hypothesis of $H_0 : \alpha = \beta = 0$ is rejected, then time series has a periodic component. Since w_k is not known, standard F-test cannot be used. Instead of F-test, Wei (2006) defines the following test statistic

Table 6 Five highest periods with their related statistics

i	$I_n(w_i)$	V_i	Period	5% critical value	Decision
1	1845.560	0.2509	12.000	0.0633	Significant
2	902.771	0.1630	240.00	0.0633	Significant
3	295.522	0.0637	12.632	0.0633	Insignificant
4	219.897	0.0507	11.429	0.0633	Insignificant
5	171.776	0.0942	21.8118	0.0633	Insignificant

$$V = I_n(w_{(1)}) \left[\sum_{k=1}^m I_n(w_k) \right]^{-1}, \quad (7)$$

where $I_n(w_{(1)})$ is the greatest periodogram value and $m = [n/2]$. Critical values of the test statistics are calculated by $c_\alpha = 1/(\alpha/m)^{1/(m-1)}$. Here, α represents the significance level. If $V_i > C_\alpha$, then null hypothesis is rejected. If the time series has periodic components,

Table 7 Parameter estimation of harmonic regression model given in Eq. (11)

Variable	Degrees of freedom	Parameter estimation	Standard error	<i>t</i> value	Pr> t
μ	1	15.04829	0.28695	52.44	<0.0001
A_1	1	2.65298	0.40581	6.54	<0.0001
B_1	1	2.88814	0.40581	7.12	<0.0001
A_2	1	− 2.66487	0.40581	− 6.57	<0.0001
B_2	1	0.64926	0.40581	1.60	0.1110

say p_1 and p_2 , the harmonic regression model with periodicity can be defined as

$$Y_t = \mu + A_1 \cos\left(\frac{2\pi t}{p_1}\right) + B_1 \sin\left(\frac{2\pi t}{p_1}\right) + A_2 \cos\left(\frac{2\pi t}{p_2}\right) + B_2 \sin\left(\frac{2\pi t}{p_2}\right) e_t, t = 1, 2, \dots, n. \quad (8)$$

The model given in Eq. (8) considers the periodicities of the time series. Thus, it is more appropriate to use such model to increase the accuracy of the forecasts.

The data used in this study were taken from the website of Airparif, a non-profit organization founded in 1979 (AIRPARIF). Airparif is accredited by the Ministry of Environment to monitor air quality in Paris and in the Ile de France region (the capital city region). As of now, measurements are made for a large number of pollutants at a total of 62 stations. PM_{2.5} measurements made at 16 stations in total since 1999 are published on the website of the organization. It is seen that PM_{2.5} measurements were obtained at more regular intervals for 4 of these 16 stations. These 4 stations are GEN (Gennevilliers), BOB (Bobigny), AUT (Autoroute A1, Saint-Denis), and VITRY (Vitry-sur-Seine). However, it was observed that PM_{2.5} measurements obtained from AUT station were above the general average obtained from other stations and had more trends. Therefore, only PM_{2.5} measurements obtained from GEN, BOB, and

VITRY stations were used in this study. Daily measurements were converted to a monthly frequency, and the period between January 2000 and December 2019 was considered. Time series plot of the data with its autocorrelation and partial autocorrelation functions is given in Fig. 1.

Results and discussion

The time series and Fig. 1 indicate that there exists a difference in PM_{2.5} concentrations from season to season. Especially it can be seen that emission rates are lower during summer and concentration level in winter is higher. Table 1 shows that there exist statistically significant differences between the months.

In order to determine the appropriate time series model for the data, different ARIMA models are considered. The model with the smallest AIC (Akaike information criteria) is chosen. We consider the following model

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + e_t, t = 1, 2, \dots, n, \quad (9)$$

where $e_t \sim WN(0, \sigma^2)$. The AIC values with the parameter estimation of σ^2 are given in Table 2.

According to Table 2, the best model for the monthly average PM_{2.5} data is

Table 8 Parameter estimation of harmonic regression model given in Eq. (12)

Variable	Degrees of freedom	Parameter estimation	Standard error	<i>t</i> value	Pr> t
μ	1	15.04829	0.28790	52.27	<0.0001
A_1	1	2.65298	0.40751	6.52	<0.0001
B_1	1	2.88814	0.40751	7.09	<0.0001
A_2	1	− 2.66487	0.40751	− 6.55	<0.0001

Fig. 3 Predictions of Harmonic Regression (Black) and observations (Blue)

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_5 Y_{t-5} + \alpha_{12} Y_{t-12} + e_t, t = 1, 2, \dots, n. \quad (10)$$

We test the unit root of the time series by employing augmented Dickey-Fuller (ADF) test. The result of the test is given in Table 3.

According to the results of ADF, test time series of $PM_{2.5}$ is stationary. In the next step, we give the parameter estimation of ARIMA model given in Eq. (10).

In Table 4, standard t test is used since the underlying data is stationary. Also, it can be seen from the table that all the parameters of ARIMA model are significant. We use ARIMA model to forecast the values of $PM_{2.5}$. Values of the observations with the forecasts are given in Table 5.

Prediction of the last 120 values with their observations are graphed in Fig. 2.

Table 9 Forecasting results of harmonic regression model

Period	Forecast	Observation
2020:01	16.1260	14.0541
2020:02	16.2148	7.6619
2020:03	15.2798	12.0619
2020:04	13.5527	13.2102
2020:05	11.5527	8.9488
2020:06	9.7632	7.4387
2020:07	8.6864	6.4281

Moreover, error sum of squares of the predictions are calculated as:

$$\sum_{i=121}^{240} (y_i - \hat{y}_{i,a})^2 = 2574.27412 \text{ and } \frac{1}{120} \sum_{i=121}^{240} (y_i - \hat{y}_{i,a})^2 = 21.4523.$$

Here, y_i represents the observations, and $\hat{y}_{i,a}$ represents the predicted values by ARIMA.

Since the data may contain periodic components, unit root test based on periodograms is conducted. To test the stationarity, the following values are calculated:

$$I_n(w_1) = 902.77, \hat{\sigma}_n^2 = 19.99638 \text{ and } t_n(w_1) = 0.030941 < c_{0.01} = 0.034818.$$

The investigated time series is stationary according to periodogram-based unit root test since the test statistics $t_n(w_1)$ is less than the 1% significance level $c_{0.01}$. The data is stationary; so we can investigate the possibility of existence of periodicity. To do this, the highest five periodogram values with their V_i statistics are given in Table 6.

According to Table 6, only periods of 12 months and 240 months are statistically significant. We can conclude that monthly $PM_{2.5}$ data has 2 periods of 12 and 240 months. By using the periods, we will estimate the harmonic regression model given in Eq. (8) as

$$Y_t = \mu + A_1 \cos\left(\frac{2\pi t}{12}\right) + B_1 \sin\left(\frac{2\pi t}{12}\right) + A_2 \cos\left(\frac{2\pi t}{240}\right) + B_2 \sin\left(\frac{2\pi t}{240}\right) + e_t, t = 1, 2, \dots, n. \quad (11)$$

Parameter estimation of the harmonic regression model given in Eq. (11) is given in Table 7.

According to Table 7, the parameter B_2 is not statistically significant. For that reason, we re-estimate the model given in Eq. (11) as

$$Y_t = \mu + A_1 \cos\left(\frac{2\pi t}{12}\right) + B_1 \sin\left(\frac{2\pi t}{12}\right) + A_2 \cos\left(\frac{2\pi t}{240}\right) + e_t, t = 1, 2, \dots, n. \quad (12)$$

The parameter estimations with significance level of model the model given in Eq. (12) is given in Table 8.

Table 8 indicates that all the parameters are significant. Estimated harmonic regression mode to forecast monthly $PM_{2.5}$ is

$$Y_t = 15.4829 + 2.65298 \cos\left(\frac{2\pi t}{12}\right) + 2.88814 \sin\left(\frac{2\pi t}{12}\right) - 2.66487 \cos\left(\frac{2\pi t}{240}\right) + e_t, t = 1, 2, \dots, n. \quad (13)$$

Graphs of the predicted values with the last 120 observations are given in Fig. 3.

Error sum of squares of the predictions are calculated as

$$\sum_{i=121}^{240} (y_i - \hat{y}_{i,h})^2 = 2520.31634 \text{ and } \frac{1}{120} \sum_{i=121}^{240} (y_i - \hat{y}_{i,h})^2 = 21.0026.$$

Here, y_i represents the observations, and $\hat{y}_{i,h}$ represents the predicted values by harmonic regression model, and $\hat{y}_{i,a}$ represents the predicted values by ARIMA model. Also, it should be noted that

$$\sum_{i=121}^{240} (y_i - \hat{y}_{i,h})^2 = 2520.31634 < 2574.27412 = \sum_{i=121}^{240} (y_i - \hat{y}_{i,a})^2.$$

The above inequality shows that error sum of squares calculated by using harmonic regression is less than the calculated error sum of squares of ARIMA. Although there exists a slight difference, we show that harmonic regression has better performance on predicting monthly $PM_{2.5}$ emission of Paris. The forecasted values by using harmonic regression are given in Table 9.

Error sum of squares is calculated for the forecasts. The calculations are as follows

$$\sum_{i=241}^{247} (y_i - \hat{y}_{i,h})^2 = 105.214669 < 218.89718 = \sum_{i=241}^{247} (y_i - \hat{y}_{i,a})^2.$$

These errors are calculated based on the test set. Here, we did not use the last 7 observations to estimate the model. As the errors suggest, the harmonic regression performs better than the ARIMA model.

Conclusion

The main goal of this study is to model and forecast monthly $PM_{2.5}$ emissions solely obtained information from itself. Instead of using multivariate time series, firstly, we have started our analysis with unit root testing. ADF test is used, and the test result showed that the investigated data was stationary. Secondly, proper time series model of ARIMA is chosen by minimizing the AIC. Thirdly, periodogram-based unit root test is conducted because of the possible hidden periodic structure of the time series. According to the periodogram-based unit root test, the time series is stationary. Thus, we can investigate the possible periodic structure. In order to investigate the periodic structure, periodograms of the time series are calculated, and it is seen that 12 months and 240 months are statistically significant. In the last step, harmonic regression is fitted by using the periodic structure of the time series.

We showed the power of univariate time series to model air pollution data. Mostly used time series methods in such cases are ARIMA type models. Although it is a very useful and powerful method, it tends to revert to the mean of the data when it is stationary. It gives consistent results for the short-term periods, but in the long run, it is not the case. Especially when the data contains periodicity, it may perform poorly. For that reason, we introduce harmonic regression model and showed that it has better performance in the sample and out of sample forecasting. Also, it can be interesting to compare the harmonic regression and ARIMA with other forecasting tools such as machine learning and deep

learning algorithms as in Ashrafzadeh et al. (2020) and Ashrafzadeh et al. (2021) in pollutant forecasting. Moreover, by utilizing such algorithms, the effects of external variables such as weather conditions, other pollutants, energy consumption, and industrial outputs can be investigated and can be compared with the harmonic regression.

We have shown that $PM_{2.5}$ has two periodicities of 1 year and 20 years. One-year periodicity corresponds to the summer months. As the data and ANOVA suggest, there exists significant difference between the means of each month. Also, it is seen that on summer, monthly emission achieves its lower value. As the literature suggest, wood burning is one of the main sources of the emission; this may be the reason of lower emission on summer. Figure 1 indicates a decreasing trend in the data. Twenty-year cycles correspond to this trend. One of the main factors of air pollutant is road transport (Angelevska et al. (2021), Abdulrazzaq et al. (2020)). As technology on car engines get advanced, they produce less emissions. Also, new rules and regulations on emission restrictions cause decreasing on the emission. Thus, this study reveals that these rules and regulations cause improvements in air pollutions within 20 years.

Acknowledgements We thank to the associate editor and the anonymous reviewers for their valuable suggestions and corrections.

Author contribution YA: Design of methodology. Application of statistical, mathematical and computational techniques to analyze study data. EG: Formulation of overarching research goals and aims. Management activities to produce metadata. KDU: Formulation of overarching research goals and aims. Conducting a research and investigation process. Preparation and presentation of the published work. MEY: Verification of the overall reproducibility, experiments and other research outputs. Preparation and presentation of the published work.

Availability of data and materials The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interest The authors declare no competing interests.

References

- Abdulrazzaq, L. R., Abdulkareem, M. N., Yazid, M. R. M., Borhan, M. N., & Mahdi, M. S. (2020). Traffic congestion: Shift from private car to public transportation. *Civil Engineering Journal*, 6(8), 1547–1554.
- Akdi, Y., & Dickey, D. A. (1998). Periodograms of unit root time series: Distributions and tests. *Communications in Statistics-Theory and Methods*, 27(1), 69–87.
- Akdi, Y., Okkaoğlu, Y., Gölveren, E., & Yücel, M. E. (2020a). Estimation and forecasting of PM 10 air pollution in Ankara via time series and harmonic regressions. *International Journal of Environmental Science and Technology*, 1–14.
- Akdi, Y., Varlik, S., & Berument, M. H. (2020b). Duration of global financial cycles. *Physica A: Statistical Mechanics and its Applications*, 124331.
- Akdi, Y., Gölveren, E., & Okkaoğlu, Y. (2020c). Daily electrical energy consumption: Periodicity, harmonic regression method and forecasting. *Energy*, 191, 116524.
- Akdi, Y., & Ünlü, K. D. (2020). Periodicity in precipitation and temperature for monthly data of Turkey. *Theoretical and Applied Climatology*. <https://doi.org/10.1007/s00704-020-03459-y>
- Anderson, J. O., Thundiyil, J. G., & Stolbach, A. (2012). Clearing the air: A review of the effects of particulate matter air pollution on human health. *Journal of Medical Toxicology*, 8, 166–175.
- Angelevska, B., Atanasova, V., & Andreevski, I. (2021). Urban air quality guidance based on measures categorization in road transport. *Civil Engineering Journal*, 7(2), 253–267.
- Ashrafzadeh, A., Kişi, O., Aghelpour, P., Biazar, S. M., & Masouleh, M. A. (2020). Comparative study of time series models, support vector machines, and GMDH in forecasting long-term evapotranspiration rates in northern Iran. *Journal of Irrigation and Drainage Engineering*, 146(6), 04020010.
- Ashrafzadeh, A., Kişi, O., Aghelpour, P., Mostafa Biazar, S., & Askarizad Masouleh, M. (2021). Closure to “comparative study of time series models, support vector machines, and gmdh in forecasting long-term evapotranspiration rates in northern Iran” by Afshin Ashrafzadeh, Ozgur Kişi, Pouya Aghelpour, Seyed Mostafa Biazar, and Mohammadreza Askarizad Masouleh. *Journal of Irrigation and Drainage Engineering*, 147(6), 07021006.
- AIRPARIF. “Air quality”. <https://www.airparif.asso.fr/en/telechargement/telechargement-polluant>. Accessed on 14 August 2020.
- Apte, J. S., Marshall, J. D., Cohen, A. J., & Brauer, M. (2015). Addressing global mortality from ambient PM_{2.5}. *Environmental Science & Technology*, 49, 8057–8066.
- Bai, Y., Zeng, B., Li, C., & Zhang, J. (2019a). An ensemble long short-term memory neural network for hourly PM_{2.5} concentration forecasting. *Chemosphere*, 222, 286–294.
- Bai, Y., Li, Y., Zeng, B., Li, C., & Zhang, J. (2019b). Hourly PM_{2.5} concentration forecast using stacked autoencoder model with emphasis on seasonality. *Journal of Cleaner Production*, 224, 739–750.
- Biancofiore, F., Busilacchio, M., Verdecchia, M., Tomassetti, B., Aruffo, E., Bianco, S., ... & Di Carlo, P. (2017). Recursive neural network model for analysis and forecast of

- PM10 and PM2. 5. Atmospheric Pollution Research, 8(4), 652–659.
- Breyse, P. N., Delfino, R. J., Dominici, F., Elder, A. C. P., Frampton, M. W., Froines, J. R., et al. (2013). US EPA particulate matter research centers: Summary of research results for 2005–2011. *Air Quality, Atmosphere & Health*, 6, 333–355.
- Brook, R. D., Newby, D. E., & Rajagopalan, S. (2017). The global threat of outdoor ambient air pollution to cardiovascular health: Time for intervention. *JAMA Cardiology*, 2(4), 353–354.
- Brockwell, P. J., & Davis, R. A. (1987). *Time series: Theory and methods*. Springer-Verlag.
- Brunekeef, B., & Holgate, S. T. (2002). Air pollution and health. *The Lancet*, 360(9341), 1233–1242.
- Chang, F. J., Chang, L. C., Kang, C. C., Wang, Y. S., & Huang, A. (2020a) Explore spatio-temporal PM2. 5 features in northern Taiwan using machine learning techniques. *Science of The Total Environment*, 139656.
- Chang, Y. S., Abimannan, S., Chiao, H. T., Lin, C. Y., & Huang, Y. P. (2020b) An ensemble learning based hybrid model and framework for air pollution forecasting. *Environmental Science and Pollution Research*, 1–14.
- Cheng, Y., Zhang, H., Liu, Z., Chen, L., & Wang, P. (2019). Hybrid algorithm for short-term forecasting of PM2. 5 in China. *Atmospheric Environment*, 200, 264–279.
- Du, J., Qiao, F., & Yu, L. (2019). Temporal characteristics and forecasting of PM2. 5 concentration based on historical data in Houston, USA. *Resources, Conservation and Recycling*, 147, 145–156.
- Dockery, D. W. (2009). Health effects of particulate air pollution. *Annals of Epidemiology*, 19(4), 257–263.
- Dockery, D. W., & Pope, C. A. (1994). Acute respiratory effects of particulate air pollution. *Annual Review of Public Health*, 15, 107–132.
- Dockery, D. W., Pope, C. A., Xu, X., Spengler, J. D., Ware, J. H., Fay, M. E., et al. (1993) An association between air pollution and mortality in six U.S. cities. *The New England Journal of Medicine*, 329(24), 1753–1759.
- Evans, K. A., Halterman, J. S., Hopke, P. K., Fagnano, M., & Rich, D. Q. (2014). Increased ultrafine particles and carbon monoxide concentrations are associated with asthma exacerbation among urban children. *Environmental Research*, 129, 11–19.
- Fuller, W. A. (1996). *Introduction to statistical time series*. Wiley.
- Franceschi, F., Cobo, M., & Figueredo, M. (2018) Discovering relationships and forecasting PM10 and PM2. 5 concentrations in Bogotá, Colombia, using artificial neural networks, principal component analysis, and k-means clustering. *Atmospheric Pollution Research*, 9(5), 912–922.
- Gasana, J., Dillikar, D., Mendy, A., Forno, E., & Vieira, E. R. (2012). Motor vehicle air pollution and asthma in children: A meta-analysis. *Environmental Research*, 117, 36–45.
- Golly, B., Waked, A., Weber, S., Samake, A., Jacob, V., Conil, S., & Besombes, J. L. (2019). Organic markers and OC source apportionment for seasonal variations of PM2. 5 at 5 rural sites in France. *Atmospheric Environment*, 198, 142–157.
- Gibergans Bàguena, J., Hervada Sala, C., & Jarauta Bragulat, E. (2020). The quality of urban air in Barcelona: A new approach applying compositional data analysis methods. *Emerging Science Journal*, 4(2), 113–121.
- Guaita, R., Pichiule, M., Mate, T., Linares, C., & Diaz, J. (2011) Short-term impact of particulate matter (PM2.5) on respiratory mortality in Madrid. *International Journal of Environmental Health Research*, 21(4), 260–274.
- Hamra, G. B., Guha, N., Cohen, A., Laden, F., Raaschou-Nielsen, O., Samet, J. M., et al. (2014). Outdoor particulate matter exposure and lung cancer: A systematic review and meta-analysis. *Environmental Health Perspectives*, 122(9), 906–911.
- Han, X., Liu, Y., Gao, H., Ma, J., Mao, X., Wang, Y., et al. (2017). Forecasting PM2.5 induced male lung cancer morbidity in China using satellite retrieved PM2.5 and spatial analysis. *Science of the Total Environment*, 607–608, 1009–1017.
- Hadley, O. L. (2017). Background PM2. 5 source apportionment in the remote Northwestern United States. *Atmospheric Environment*, 167, 298–308.
- Hart, J. E., Liao, X., Hong, B., Puett, R. C., Yanosky, J. D. Suh, H., et al. (2015) The association of long-term exposure to PM2.5 on all-cause mortality in the nurses’ health study and the impact of measurement-error correction. *Environmental Health*, 14(38), 1–9.
- Hoek, G., Krishnan, R. M., Beelen, R., Peters, A., Ostro, B., Brunekreef, B., et al. (2013). Long-term air pollution exposure and cardio-respiratory mortality: A review. *Environmental Health*, 12(43), 1–15.
- Kim, K. H., Kabir, E., & Kabir, S. (2015). A review on the human health impact of airborne particulate matter. *Environment International*, 74, 136–143.
- Kow, P. Y., Wang, Y. S., Zhou, Y., Kao, I. F., Issermann, M., Chang, L. C., & Chang, F. J. (2020) Seamless integration of convolutional and back-propagation neural networks for regional multi-step-ahead PM2. 5 forecasting. *Journal of Cleaner Production*, 121285.
- Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., & Pozzer, A. (2015). The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature*, 525(7569), 367–371.
- Lin, S., Munsie, J. P., Hwang, S. A., Fitzgerald, E., & Cayo, M. R. (2002). Childhood asthma hospitalization and residential exposure to state route traffic. *Environmental Research*, 88(2), 73–81.
- Lippmann, M. (2014) Toxicological and epidemiological studies of cardiovascular effects of ambient air fine particulate matter (PM2.5) and its chemical components: Coherence and public health implications. *Critical Reviews in Toxicology*, 44(4), 299–347.
- Liu, H. Y., Dunea, D. N. Iordache, S., & Pohoata, A. (2018). A review of airborne particulate matter effects on young children’s respiratory symptoms and diseases. *Atmosphere*, 9(4).
- Loftus, C., Yost, M., Sampson, P., Arias, G., Torres, E., Vasquez, V. B., et al. (2015). Regional PM2.5 and asthma morbidity in an agricultural community: A panel study. *Environmental Research*, 136, 505–512.
- Lopez-Restrepo, S., Yarce, A., Pinel, N., Quintero, O. L., Segers, A., & Heemink, A. W. (2020) Forecasting PM10 and PM2. 5 in the Aburrá Valley (Medellín, Colombia) via EnKF based data assimilation. *Atmospheric Environment*, 117507.

- Lv, B., Cobourn, W. G., & Bai, Y. (2016). Development of nonlinear empirical models to forecast daily PM_{2.5} and ozone levels in three large Chinese cities. *Atmospheric Environment*, 147, 209–223.
- Maji, K. J., Dikshit, A. K., Arora, M., & Deshpande, A. (2018). Estimating premature mortality attributable to PM_{2.5} exposure and benefit of air pollution control policies in China for 2020. *Science of the Total Environment*, 612, 683–693.
- Maji, K. J., Dikshit, A. K., & Deshpande, A. (2017). Disability-adjusted life years and economic cost assessment of the health effects related to PM_{2.5} and PM₁₀ pollution in Mumbai and Delhi, in India from 1991 to 2015. *Environmental Science and Pollution Research*, 24, 4709–4730.
- Mannucci, P. M., & Franchini, M. (2017). Health effects of ambient air pollution in developing countries. *International Journal of Environmental Research and Public Health*, 14(9), 1048.
- Moisan, S., Herrera, R., & Clements, A. (2018). A dynamic multiple equation approach for forecasting PM_{2.5} pollution in Santiago, Chile. *International Journal of Forecasting*, 34(4), 566–581.
- Okkaoglu, Y., Akdi, Y., & Ünlü, K. D. (2020). Daily PM₁₀, periodicity and harmonic regression model: The case of London. *Atmospheric Environment*, 117755.
- Orru, H., Maasikmets, M., Lai, T., Tamm, T., Kaasik, M., Kimmel, V., et al. (2011). Health impacts of particulate matter in five major Estonian towns: Main Sources of Exposure and Local Differences. *Air Quality, Atmosphere & Health*, 4, 247–258.
- Ostro, B. D., Lipsett, M. J., & Das, R. (1998). Particulate Matter and Asthma: A quantitative assessment of the current evidence. *Applied Occupational and Environmental Hygiene*, 13(6), 453–460.
- Peled, R. (2011). Air pollution exposure: Who is at high risk? *Atmospheric Environment*, 45(10), 1781–1785.
- Perez, P., Menares, C., & Ramírez, C. (2020). PM_{2.5} forecasting in Coyhaique, the most polluted city in the Americas. *Urban Climate*, 32, 100608.
- Perez, P., & Gramsch, E. (2016). Forecasting hourly PM_{2.5} in Santiago de Chile with emphasis on night episodes. *Atmospheric Environment*, 124, 22–27.
- Pope, C. A., Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K., et al. (2002). Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA*, 287(9), 1132–1141.
- Pope, C. A., & Dockery, D. W. (2006). Health effects of fine particulate air pollution: Lines that connect. *Journal of the Air & Waste Management Association*, 56(6), 709–742.
- Pui, D. Y. H., Chen, S. C., & Zuo, Z. (2014). PM_{2.5} in China: Measurements, sources, visibility and health effects, and mitigation. *Particuology*, 13, 1–26.
- Qi, Y., Li, Q., Karimian, H., & Liu, D. (2019). A hybrid model for spatiotemporal forecasting of PM_{2.5} based on graph convolutional neural network and long short-term memory. *Science of the Total Environment*, 664, 1–10.
- Querol, X., Alastuey, A., Rodriguez, S., Plana, F., Ruiz, C. R., Cots, N., ... & Puig, O. (2001). PM₁₀ and PM_{2.5} source apportionment in the Barcelona Metropolitan area, Catalonia, Spain. *Atmospheric Environment*, 35(36), 6407–6419.
- Rao, M., D'Elia, I., & Piersanti, A. (2018). An uncertainty quantification of PM_{2.5} emissions from residential wood combustion in Italy. *Atmospheric Pollution Research*, 9(3), 526–533.
- Rojas-Bracho, L., Suh, H. H., & Koutrakis, P. (2000). Relationships among personal, indoor, and outdoor fine and coarse particle concentrations for individuals with COPD. *Journal of Exposure Analysis and Environmental Epidemiology*, 10, 294–306.
- Saxon, A., & Diaz-Sanchez, D. (2000). Diesel exhaust as a model xenobiotic in allergic inflammation. *Immunopharmacology*, 48(3), 325–327.
- Schikowski, T., Mills, I. C., Anderson, H. R., Cohen, A., Hansell, A., Kauffmann, F., et al. (2014). Ambient air pollution: A cause of COPD? *European Respiratory Journal*, 43, 250–263.
- Schwartz, J., Dockery, D. W., & Neas, L. M. (1996). Is daily mortality associated specifically with fine particles? *Journal of the Air & Waste Management Association*, 46(10), 927–939.
- Shou, Y., Huang, Y., Zhu, X., Liu, C., Hu, Y., & Wang, H. (2019). A review of the possible associations between ambient PM_{2.5} exposures and the development of Alzheimer's disease. *Ecotoxicology and Environmental Safety*, 174, 344–352.
- Stanek, L. W., Sacks, J. D., Dutton, S. J., & Dubois, J. J. B. (2011). Attributing health effects to apportioned components and sources of particulate matter: An evaluation of collective results. *Atmospheric Environment*, 45(32), 5655–5663.
- Shang, Z., Deng, T., He, J., & Duan, X. (2019). A novel model for hourly PM_{2.5} concentration prediction based on CART and EELM. *Science of the Total Environment*, 651, 3043–3052.
- Shahid, I., Kistler, M., Mukhtar, A., Ghauri, B. M., Ramirez-Santa Cruz, C., Bauer, H., & Puxbaum, H. (2016). Chemical characterization and mass closure of PM₁₀ and PM_{2.5} at an urban site in Karachi-Pakistan. *Atmospheric Environment*, 128, 114–123.
- Turner, M. C., Krewski, D., Pope, C. A., Chen, Y., Gapstur, S. M., & Thun, M. J. (2011). Long-term ambient fine particulate matter air pollution and lung cancer in a large cohort of never smokers. *American Journal of Respiratory and Critical Care Medicine*, 184(12), 1374–1381.
- Wang, P., Zhang, G., Chen, F., & He, Y. (2019). A hybrid-wavelet model applied for forecasting PM_{2.5} concentrations in Taiyuan city, China. *Atmospheric Pollution Research*, 10(6), 1884–1894.
- Wei, W. W. S. (2006). *Time series analysis: Univariate and multivariate methods*. Pearson Education.
- Weiland, S. K., Mundt, K. A., Diplstat, A. R., & Keil, U. (1994). Self-reported wheezing and allergic rhinitis in children and traffic density on street of residence. *Annals of Epidemiology*, 4(3), 243–247.
- WHO. World Health Organization, (2020a) "Air pollution". https://www.who.int/health-topics/air-pollution#tab=tab_1. Accessed on 21 September 2020.
- WHO. Regional Office for Europe, (2020b) "Air quality". <https://www.euro.who.int/en/health-topics/environment-and-health/air-quality>. Accessed on 23 September 2020.

- Xu, X., Tong, T., Zhang, W., & Meng, L. (2020) Fine-grained prediction of PM_{2.5} concentration based on multisource data and deep learning. *Atmospheric Pollution Research*. <https://doi.org/10.1016/j.apr.2020.06.032>
- Yuan, W., Wang, K., Bo, X., Tang, L., & Wu, J. (2019) A novel multi-factor & multi-scale method for PM_{2.5} concentration forecasting. *Environmental Pollution*, 255, 113187.
- Zhang, L., Lin, J., Qiu, R., Hu, X., Zhang, H., Chen, Q., & Wang, J. (2018). Trend analysis and forecast of PM_{2.5} in Fuzhou, China using the ARIMA model. *Ecological Indicators*, 95, 702–710.
- Zhou, Y., Chang, F. J., Chang, L. C., Kao, I. F., Wang, Y. S., & Kang, C. C. (2019). Multi-output support vector machine for regional multi-step-ahead PM_{2.5} forecasting. *Science of the Total Environment*, 651, 230–240.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.