Estimation and forecasting of PM_{10} air pollution in Ankara via time series and harmonic regressions

Y. Akdi[1] · Y. Okkaoğlu[2] · E. Gölveren[3] · M. E. Yücel[4]

Abstract

In this study, monthly particulate matter (PM_{10}) values in Ankara (39.9334° N, 32.8597° E) from January 1993 to December 2017 are examined. The PM_{10} are those thoracic particles whose aerodynamic diameter is less than 10 μm (micrometers), and it is of critical health importance due to the penetrability to the lower airways. As an alternative to classical unit root tests, a unit root test primarily based on periodograms is introduced owing to its advantages over alternatives. After examining the stationarity of the series through periodogram-based test as well as its standard rivals, periodic components in the series are examined and it is observed that the series has both periodic and seasonal components. These components are modeled, using the inherent dynamics of a time series alone, within a trigonometric harmonic regression setup, eventually yielding the forecast values for 2018 that turns out to be superior to those obtained by means of ARIMA (autoregressive integrated moving average). This is a striking result since the modeling framework requires no assumptions, no parameter estimations except for the variance of the white noise series, no simulations of the power of tests, no adjustments of test statistics with respect to sample size and no preliminary work as to independent variable which is simply time, i.e., the period of forecast.

Keywords Time series analysis · Periodogram · Harmonic regression · Air pollution · PM_{10} concentration · Forecasting

JEL Classification C13 · C53 · C8

Introduction

Air pollution can be defined as the phenomenon that various substances that need to be found in the air are out of the specified limits and the substances that should not be in the air are found at a hazardous rate to humans, plants, animals

✉ Y. Okkaoğlu
yo2u19@soton.ac.uk

1   Department of Statistics, Faculty of Science, University of Ankara, Ankara, Turkey

2   Department of Mathematical Sciences, Faculty of Social Sciences, University of Southampton, Southampton, UK

3   Department of Econometrics, Faculty of Economics and Administrative Sciences, Ataturk University, Erzurum, Turkey

4   Department of Economics, Faculty of Economics, Administrative and Social Sciences, Ihsan Dogramaci Bilkent University, Ankara, Turkey

and environment (Cavkaytar et al. 2013). Along with various natural factors, air pollution increasing in parallel with human factors related to population growth, technological development and industrialization, has long been a global concern. Among all factors, increasing population density is one of the key human factors that cause air pollution without doubt. In that, the impacts of the natural pace of population increase were visibly augmented by the migration from rural to urban areas in Turkey especially since 1980. According to TurkStat, while Turkey's population was 43.9 million in 1980, this figure reached 63.2 million in 2000 and 80.8 million in 2017. Not surprisingly, the most dramatic change in Turkey's demographics has been the rate of urbanization. While the urban–rural divide was 24.2–75.8% in 1927, today this rate is reversed to 92.5–7.5% as of 2017.

A similar tendency is seen in metropolitan areas, as well. For instance, the population in Ankara reached from 3,889,199 in 2000 to 4,771,716 in 2010 and 5,445,026 in 2017, where 94.3% of city population live in the centrum (TurkStat 2018). Such an increase in population and urbanization calls for a number of environmental issues. In Ankara, the second largest

city and capital of Turkey, air pollution has been a constant problem. The increase in the number of vehicles in traffic and their consumption of fossil fuels induced largely by population growth are its most important sources. Although the use of natural gas, removal of leaded gasoline consumption, emphasis on the use of natural gas-powered public transport vehicles initiated in 1988 slowed this increase, susceptibility of Ankara to air pollution in winters has not yet disappeared.

A brief look at the contaminants that threaten human health underlines carbon monoxide (CO), particulate matter ($PM_{2.5}$ and PM_{10}), sulfur dioxide (SO_2), nitrogen oxides (NO_x) and ozone (O_3) as noted by Li and Shi (2016). Particulate matter (PM) affects human health more than other pollutants (WHO 2018) and is defined as a mixture of solid particles and liquid droplets, hanging in the atmosphere. The PM_{10} represents thoracic particles whose aerodynamic diameter is less than 10 μm and can penetrate to the lower airways. The PM_{10} pollutants contain airborne particles due to the use of carbon-containing fuels, cigarette consumption and various industrial wastes, caused by humans, as well as particles formed due to the naturally occurring volcanic gases, seawater vapor, dust storms and forest fires which are mixed to atmosphere. The PM_{10} causes many lung diseases, especially asthma, and cardiovascular diseases (Cavkaytar et al. 2013; Bayram and Dikensoy 2006). The World Health Organization (WHO) has set the annual average PM_{10} amount standard as 20 μg/m^3 and the 24-hour average amount of PM_{10} standard as 50 μg/m^3 in the 2005 Air Quality Directive WHO (2018). The European Parliament and Council has determined the annual average PM_{10} limit as 40 μg/m^3 and the 24-hour average PM_{10} limit as 50 μg/m^3 under the directive 2008/50/EC of May 21, 2008, and that these values can be exceeded maximum 35 times per year European Commission (2018). In Turkey, according to Chamner of Environmental Engineers the amount of 24-hour average PM_{10} limit is determined as 80 μg/m^3 in 2016 and this value has been updated as 60 μg/m^3 in 2018. According to the Air Pollution Report for 2017, the PM_{10} concentrations recorded for Ankara in 2017 exceeded the upper bounds for more days than in 2016. As the Chamber of Environmental Engineers (2018) noted, such an incidence may be pointing at potential hazards to human health.

Note that, with regard to different particulate matters, $PM_{2.5}$ could also have been considered in our analysis. $PM_{2.5}$ are those particulate matters that have a diameter of less than 2.5 μm (micrometers), which is about 3% of the diameter of human hair. Since they are even smaller than their PM_{10} counterparts, they are also called fine particles and have a higher likelihood to float longer in air. Nevertheless, our analysis could not consider $PM_{2.5}$ due to the inavailability of data on our side. Another important venue could be the consideration of the type of source of the PMs, i.e., a separate analysis of PMs from local sources (road traffic, industries, firewood burning, etc.,) and from distant sources (background PMs) could have shed more light to our understanding, basically since the variation in the concentration of PMs throughout the day and over the months of the year is very different and depends heavily on the source under study. However, we could not entail these in our analysis due to the inavailability of data with regard to sources of PM_{10}.

Against this background, in this study, we propose an appropriate time series model for the monthly PM_{10} amounts observed in Ankara from January 1993 to December 2017 and generate a set of forecasts for future. The monthly PM_{10} measurements (μg/m^3) for Ankara from January 1993 to December 2017 are used. The measurements between January 2011 and October 1993 are taken from TurkStat, while the measurements between December 2017 and November 2011 are taken from Air Pollution Monitoring Network of the Ministry of Environment and Urban Affairs. Data are compiled as monthly averages of observations taken from eight stations located in Sıhhiye (39.928105° N, 32.852785° E), Bahçelievler (39.928741° N, 32.823128° E), Dikmen (39.877900° N, 32.834940° E), Cebeci (39.931606° N, 32.877861° E), Demetevler (39.964339° N, 32.780970° E), Keçiören (39.996689° N, 32.811290° E), Sincan (39.966766° N, 32.575474° E) and Kayaş (39.911815° N, 32.964695° E) in Ankara, where the parentheses include the approximate locations of them.

Considering the inherent features of seasonality and periodicity in data, we maintain a periodogram-based investigation. Our forecasting models are obtained through harmonic regression techniques which consider periodicity besides classical time series ingredients. Then, the results obtained for these two models are statistically compared and the forecast values obtained according to both models are compared with the observed values for the course of 2018. These comparisons reveal a great deal of forecasting superiority in our proposed estimation framework. What distinguishes the analysis of this study from the earlier work is that our proposed methodology makes use of a single time series and extracts the information embedded in the series by eloquently handling the repeating behavior, i.e., the seasonality and periodicity. So, even in the absence of other variables, which might be of concern under another analytical scheme, our work shows that a better forecasting performance has been viable.

In the remainder of the article, a brief review of the related literature is provided in Sect. 2 and our methodology is introduced in Sect. 3. Section 4 provides the implementation of method, prior to discussions in Sect. 5.

Literature

As mentioned earlier, people of today are exposed to air pollutants beyond standards, especially in areas with high population density, and their consequent hazards to health

(Cavkaytar et al. 2013). When the literature on air pollution is examined, it can be seen that the studies are clustered around three main themes: (1) the studies intended to determine the relationship between pollutants and meteorological parameters and to determine the most important causes of air pollution, (2) the studies that demonstrate the relationship between pollutants and various health indicators and (3) the studies intended for forecasting various pollutants for the near future.

Among the studies intended to determine the relationship between pollutants and meteorological parameters and to determine the key causes of air pollution, Çiçek et al. (2004) use the stepwise regression to exhibit the relationship between SO_2, PM_{10}, NO, NO_2, CO values and meteorological parameters such as the temperature, wind speed and relative humidity for the period of November 2001–April 2002 in Sıhhiye District of Ankara. They reveal especially for March, the middle-level relation between climate elements and SO_2, PM_{10}, NO, NO_2, CO concentrations. Genç et al. (2010) employ a multiple linear regression setup for the air pollution index formed by the amounts of PM_{10} and SO_2 taken from different settlements of Ankara in 1999–2000, and they propose an air pollution index (I_c) as in Eq. (1):

$$I_c = 0.212\Delta T + 0.043P + 0.162I_{d-1} - 1.705V - 27.945, \quad (1)$$

where ΔT is the daily temperature change, P is the daily mean atmospheric pressure, I_{d-1} is the air pollution index of the previous day and V is the daily average wind speed. In order for such a regression model to be applied, the variables in the explanatory variable role must be non-random variables and the variables in the dependent variable role must be independent random variables. Also, since the observed values used are the values obtained in unit time intervals, it is more suitable to apply the time series models instead of the regression model. Nevertheless, it is still possible to use regression techniques, but firstly the data must meet requirements of the stationarity tests. Finally, Koutrakis et al. (2005) examine the relationships between $PM_{2.5}$, PM_{10}, $PM_{2.5-10}$ and meteorological variables using mixed regression models where they estimate the specific factors affecting the particle concentrations and their relative effects in Santiago, Chile, in 1989–2001. Koutrakis et al. (2005) report significant relationships between meteorological variables and particulate matter amounts, like the reduction in the amount of particulate matter on Sundays as a result of reduced traffic and other polluting activities.

Silva (2015) is another important work in that the authors developed an index for air and noise quality via aggregation of relevant measurements and presented their results in comparison with the standardized legal limits for air pollution and noise. As an aid in decision making for urban planners and various policy makers, Silva's (2015) index is of high value-added and widespread benefits. Owing to its intuitive methodology that involves the computation of a weighted linear combination of two base indices (with regard to noise and air separately), Silva (2015) provides a good description of the subject matter. The recent work by Ganguly et al. (2019) is also worthwhile with regard to measurement of air pollution along with its connection to other relevant factors. In that, the study provides not only an assessment of measurements at the urban stations in comparison with background monitoring facilities, but also the linkages of measured pollutant concentrations to seasonal factors and other pollutants' behavior. These two recent papers, in addition to those mentioned before, lay down a solid basis for understanding the problem at hand.

In the second strand, among the studies that investigate the relationship between pollutants and health indicators, to establish the relationship between the asthmatic response and the exposure of SO_2 and PM_{10}, Berktaş and Bircan (2003) consider the number of patients who admitted to emergency room (Ankara Atatürk Chest Diseases and Chest Surgery Training and Research Hospital) with complaint of asthma between January 1, 1998, and December 31, 1998, meteorological conditions (daily mean rainfall, actual pressure, relative humidity, wind speed, duration and direction as well as minimum, average and maximum daily temperatures) and SO_2 and PM_{10} concentrations in the Ankara region. Pearson and Spearman correlation tests and chi-square test were used to reveal the inherent statistical relationships. A significant correlation between the amount of PM_{10} and the number of patients who applied was found, whereas a significant low level of association between the amount of SO_2 and the number of patients who applied to emergency service due to asthma was found. The results show that even short-term exposure to low-level air pollution in Ankara increases emergency room visits for suspected asthmatic reactions. Pope and Dockery (1992) study the acute health impact of respirable particulate pollution on symptomatic and asymptomatic children during 1990–1991 in the Utah Valley. Using logistic regression analysis, a positive association between PM_{10} and respiratory symptoms was found. Ostro et al. (1999) reported a significant correlation between the PM_{10} amount and the daily mortality in the Coachella Valley by means of Poisson regressions.

The last strand of the reviewed literature is devoted to forecasting of various pollutants for the near future. Turgut and Temiz (2015) apply the Box–Jenkins methodology to the weekly PM_{10} concentration data obtained from the Ankara Sıhhiye station from January 1, 2010, to October 31, 2014, and forecast the future values of PM_{10}. The study forecasts via an ARIMA (3,0,0) specification for November 2014, December 2014 and January 2015, yet omitting the seasonal effects which have been inherent to data. Another study for Ankara Province estimates an Air Quality Health Index comprising of concentration of pollutants such as

PM$_{2.5}$, O$_3$ and NO$_2$ (Bozkurt et al. 2015). Kurt et al. (2008) and Saral (2000) employ neural networks to predict air pollution in Istanbul, Kaplan et al. (2014) use Levenberg–Marquardt learning algorithm in artificial neural networks for PM$_{10}$ and SO$_2$ estimations for Kütahya Province and Yüksek et al. (2007) estimate the SO$_2$ pollution for Sivas centrum using a backpropagation artificial neural network framework.

All in all, it is salutary that the literature is deserted neither on air pollution nor on its determinants or consequences. Nevertheless, there is a common tendency to use regression rather than time series techniques to study PM$_{10}$ pollutant in Ankara, despite the genuine time series structure of the data sets. Indeed, this very time series structure might allow researchers to extract some rich information from data even in the absence of other explanatory variables or modeling peculiarities. Even when a pure times series approach is maintained, it is not rare to observe improper or no handling of seasonality. Among practitioners, it is not rare to observe the use of standard ADF[1] (augmented Dickey–Fuller) rather than DHF[2] (Dickey, Hasza and Fuller) or HEGY[3] (Hylleberg, Engle, Granger ve Yoo) tests even when seasonality is obvious. So, we try, in what follows to avoid these pitfalls through a genuine time series approach to modeling and forecasting.

Materials and methods

The fundamental aim in the time series analysis is to forecast the future values of a series using its observed (past) values. Stationarity[4] is the most important assumption for the series

---

[1] Test procedure which proposed by Said and Dickey (1985) for unit root test in *ARIMA*($p, 1, q$) models.

[2] Seasonal unit root test which proposed by Dickey, Hasza, and Fuller (1984) and based on the symmetric least squares estimator of the parameters.

[3] Seasonal unit root test which proposed by Hylleberg et al. (1990) and considered quarterly data.

[4] A time series $\left\{ Y_t : t \in T \right\}$ is a collection of observations at unit time intervals, where T is a set of indices. Any time series can be handled as moving average (MA), autoregressive (AR), autoregressive moving average (ARMA) or seasonal processes based on how it fluctuates around its own mean. Despite the resemblance of especially autoregressive time series with regression models, they do differ in the basic assumptions. Omission of this fact may often result in misleading results while studying time series data. Assume that a given time series Y_t, $t = 1, 2, \ldots, n$, or $\left\{ Y_1, Y_2, \ldots, Y_n \right\}$, fit to the *AR*($p$) representation as in Eq. (2):

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

where e_t is a white noise series with expected value 0 and variance σ^2, that is, $e_t \sim WN\left(0, \sigma^2\right)$. The characteristic equation for Eq. (2) is written as in Eq. (3):

$$m^p - \sum_{i=1}^{p} \alpha_i m^{p-i} = 0. \quad (3)$$

to be forecastable. If the series is non-stationary, the forecasts obtained and the statistical inference about the model parameters will not be meaningful. MA series are always stationary, but AR series may not be stationary. Moreover, most of the economic series are non-stationary series. In order to make forecast by the non-stationary time series, you need to provide the stationary with the help of various transformations. There are many methods in the literature for testing the stationary of time series. But the two tests come into prominence in terms of both practicality and applicability: Dickey–Fuller test based on the distribution of the least squares estimators of the parameters and the Phillips–Perron test that used the critical values of this distribution. For these methods, the values of test statistics and p values are directly calculated by many package programs. However, the DHF method which is developed by depending on the distribution of the symmetric least squares estimator or the HEGY method is used to test the stationary of the seasonal series. In either case, an auxiliary regression model is utilized. If there is not a suspect about the periodicity[5] of the series, one of the above-mentioned tests can be used for the stationarity test. If the series contains a periodic component, the stationarity test based on periodograms can be used. Even if the series do not contain periodicity it can use, at the same time it can apply to the seasonal series (Akdi and Dickey 1999).

Periodograms are usually used to reveal hidden periodicities found in the series (Fuller 1996; Wei 2006; Brockwell and Davis 1987). Akdi and Dickey (1998) proposed a test based on periodograms. General explanations about periodograms are given below.

Periodic functions suggest trigonometric functions. So, to test whether the series contains a periodic component or not, any time series $\left\{ Y_1, Y_2, \ldots, Y_n \right\}$ can be defined by

$$Y_t = \mu + R \cos\left(wt + \phi\right) + e_t, \quad t = 1, 2, \ldots, n, \quad (4)$$

where μ, R, w and ϕ are referred to as expected value, amplitude, frequency and phase, respectively. It is necessary to estimate these parameters. Furthermore, when $w_k = 2\pi k/n$, the Fourier frequencies are obtained. Due to the characteristics of cosine functions, for $\alpha = R \cos\left(\phi\right)$ and $\beta = R \sin\left(\phi\right)$, this model can be written as

---

Footnote 4 (continued)

If all the roots, in absolute value, of Equation (3) are less than 1, the specification of Eq. (2) is said to be stationary. The case where the root is larger than 1 is not so common in practice, and the series with one of the roots has the absolute value of 1 are unit-rooted (non-stationary) series.

[5] If a real-valued function f satisfies $f(x + p) = f(x)$, then f is said to be "periodic" with a period of p. If f is periodic with p, all multiples of p are periods of f since $f(x + 2p) = f((x + p) + p) = f(x + p) = f(x)$. The smallest of these periods is the fundamental period (or briefly period) of the function.

$$Y_t = \mu + \alpha \cos\left(w_k t\right) + \beta \sin\left(w_k t\right) + e_t, \quad t = 1, 2, \ldots, n. \tag{5}$$

According to this model, if the null hypothesis $H_0 : \alpha = \beta = 0$ is rejected, it is decided that the data contain the periodic component. Standard F test can be used to test this hypothesis. However, the use of the F statistic is not significant in case w_k frequencies are not known (Wei 2006). According to this model, the least squares estimators of μ, α and β parameters are, respectively:

$$\hat{\mu} = \overline{Y}_n, a_k = \frac{2}{n} \sum_{t=1}^{n} \left(Y_t - \overline{Y}_n\right) \cos\left(w_k t\right) \text{ve}$$

$$b_k = \frac{2}{n} \sum_{t=1}^{n} \left(Y_t - \overline{Y}_n\right) \sin\left(w_k t\right), \tag{6}$$

where a_k and b_k are called as *Fourier coefficients*. Due to the characteristics of cosine functions, since

$$\sum_{t=1}^{n} \cos\left(w_k t\right) = \sum_{t=1}^{n} \sin\left(w_k t\right) = 0, \tag{7}$$

Fourier frequencies are invariant according to the mean. By means of these Fourier coefficients, the periodogram ordinate at w_k frequency of the time series is calculated as

$$I_n\left(w_k\right) = \frac{n}{2}\left(a_k^2 + b_k^2\right). \tag{8}$$

Time series are usually examined under time domain and frequency domain. While the autocorrelation function is the most important point in the time domain, the spectral density function is important in the frequency domain. There is also a transition between autocorrelation and spectral density function of series as stated in Herglotz theorem. If $f\left(w_k\right)$ is the spectral density function of the stationary time series, then the asymptotic distribution of the statistic $I_n\left(w_k\right)/f\left(w_k\right)$ becomes a chi-square distribution with 2 degree of freedom (exponential distribution with expected value of 1). That is, the probability density function of the asymptotic distribution of normalized periodograms is

$$f(x) = \begin{cases} e^{-x}, & x > 0 \\ 0 & \text{dy}. \end{cases} \tag{9}$$

Hence, periodograms can be taken as an estimator of the spectral density function. The periodograms are also used for the investigation of probable periodicities in the data (for testing the hypothesis H_0 above). For any stationary time series, the periodogram values $I_n\left(w_k\right)$ at each k frequency are calculated. The statistic V is defined by

$$V = I_n\left(w_{(1)}\right) \left[\sum_{k=1}^{m} I_n\left(w_k\right)\right]^{-1}, \tag{10}$$

where $I_n\left(w_{(1)}\right)$ is the greatest periodogram value and $m = n/2$. If there is no any periodic component in the data (under $H_0 : \alpha = \beta = 0$), then for the V statistic

$$P\left(V > c_\alpha\right) = \alpha \cong m\left(1 - c_\alpha\right)^{m-1} \tag{11}$$

can be written (Wei 2006). For any selected level of α significance, the critical value c_α is calculated by

$$c_\alpha = 1 - (\alpha/m)^{1/(m-1)}. \tag{12}$$

If $V > c_\alpha$, then the hypothesis $H_0 : \alpha = \beta = 0$ is rejected and it is concluded that the series include periodic component.

If the given time series is stationary, it was stated that the asymptotic distribution of the normalized periodogram is a chi-square distribution with 2 degree of freedom. In that case, it is written as

$$I_n\left(w_k\right)/f\left(w_k\right) \xrightarrow{D} \chi_2^2, n \to \infty \tag{13}$$

(Fuller 1996; Wei 2006; Brockwell and Davis 1987). Under the assumption that the series is not stationary in other words that it is unit-rooted, for each constant w_k, it is

$$T_n\left(w_k\right) = \frac{2\left(1 - \cos\left(w_k\right)\right)}{\hat{\sigma}_n^2} I_n\left(w_k\right) \xrightarrow{D} Z_1^2 + 3Z_2^2, n \to \infty, \tag{14}$$

where Z_1 and Z_2 are the independent variables which have the standard normal distribution and $\hat{\sigma}_n^2$ is an estimator of the variance of the error term (Akdi and Dickey 1998). Briefly, the asymptotic distribution is

$$T_n\left(w_k\right) \xrightarrow{D} \chi_1^2 + 3\chi_1^2, n \to \infty. \tag{15}$$

Again, it has been shown by Akdi and Dickey (1999) that the method is also applicable for the seasonal time series (Akdi and Dickey 1999). That is, the statistic $T_n\left(w_k\right)$ can also be used to test the stationary of the seasonal series (whether it is unit root or not). Although the asymptotic distribution is valid for each constant w_k, the frequency w_1 is usually used in the hypothesis tests. The critical values of the distribution are given by the authors.

The structure elaborated in Eq. (4) through (15) has an array of advantages for the modeler with regard to assumptions, distributions of test statistics and accuracy of numerical outcomes. In that, first, no model assumption is needed and the method is invariant to model specifications as the periodograms can be calculated without reference to model assumptions. Second, no parameter estimation is required except for the variance of the white noise series, as opposed to the standard unit root tests which require estimated model parameters first. Third, as the distribution of the statistic $T_n\left(w_k\right)$ is known under H_0 and H_a, the analytical power

function exists for the test. Fourth, the critical values of the test statistic do not depend on the sample size. Finally, since the periodograms are calculated through trigonometric transformations of data, any periodic components of data are to be captured by the method, a clear strength of the framework to yield more meaningful and accurate results eventually.

Analysis

As mentioned earlier, a chief capability of the methodology maintained in this paper is to reveal the information embedded in a time series by focusing solely on the time series itself, i.e., not explicitly referring to other variables. Such a capability on the side of modeling allows us to keep some factors outside the analysis. For instance, the prevailing meteorological conditions like wind direction, wind speed, etc., are not under our direct consideration. The bright side is that such an omission does not reduce the information content of our findings, since the impacts of these factors have already been established in the evolution of the time series over time. Equivalently, the numerous effects that seem to have been omitted are already captured as an array of periodic components in our periodogram-based analysis. Consequently, in this section the monthly PM_{10} measurements (μg/m^3) for Ankara from January 1993 to December 2017 are used. The measurements between January 2011 and October 1993 are taken from TurkStat, while the measurements between December 2017 and November 2011 are taken from Air Pollution Monitoring Network of the Ministry of Environment and Urban Affairs. Data are compiled as monthly averages of observations taken from eight stations located in Ankara. The eight missing observations from 2006 to 2007 were interpolated using the behavior of 1993:01–2006:03 (Fig. 1).

Although there is a significant reduction in the PM_{10} amounts especially after substitution of coals with natural gas in the late 1990s, the values are still above the European Union, World Health Organization and United Nations standards. When the monthly behavior of PM_{10} values is considered (Table 1), it is seen that the averages for winter months are higher than those for the others. The highest values are observed in November, December and January. The lowest value is observed in June, July and August, though not falling below international standards.

To see whether the PM_{10} values differ according to the months or not, the following one-way ANOVA (analysis of variance) model is considered:

$$y_{ij} = \mu + \alpha_i + e_{ij}, \quad i = 1, 2, \ldots, 12, j = 1, 2, \ldots, 25, \quad (16)$$

where e_{ij}'s are the independent variables which have the normal distribution. As a result of the analysis (Table 2), the null hypothesis of $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_{12} = 0$ is rejected

(F value $= 34, 33$, p value < 0.0001), i.e., the air pollution is said not to differ across months.

Time-domain approach

To determine a suitable time series model for the data, different candidates are considered and the model with the smallest AIC[6] (Akaike information criterion) statistic value (Table 3) is chosen.

Considering the results in Table 3, it can be said that the AR(13) is the most suitable model (Model I) for the data. Accordingly, the model considered is

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_{12} Y_{t-12}$$
$$+ \alpha_{13} Y_{t-13} + e_t, \quad t = 1, 2, \ldots, n, \quad (17)$$

where $e_t \sim WN(0, \sigma^2)$. The parameter estimates for this model are given in Table 4.

According to the results obtained with both the PROC ARIMA and the PROC REG, the most suitable model (Model II) for the data is

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_{12} Y_{t-12} + \alpha_{13} Y_{t-13} + e_t, \quad t = 1, 2, \ldots, n. \quad (18)$$

On the other hand, the value of the AIC statistic obtained for this model is very close to the value obtained for the AR(13) model. The parameter estimates are presented in Table 5. The results show that the model parameters are significant.

Stationarity

If the sum of the estimated values of the parameters in the model is close to 1, then this case indicates that the model is stationary. When the estimation results obtained for the Model II are considered (Table 5), it is seen that the results are different from each other such that $\hat{\alpha}_1 + \hat{\alpha}_{12} + \hat{\alpha}_{13} \cong 0.9717$ for PROC ARIMA and $\hat{\alpha}_1 + \hat{\alpha}_{12} + \hat{\alpha}_{13} \cong 0.8361$ for PROC REG. In order to specify exactly whether the assumption of stationarity holds, the unit root test has been performed. The results of the ADF and PP[7] (Phillips–Perron) unit root test with 13 the maximum delay length are presented in Table 6. The results show that the series is stationary.

After determining that the series is stationary, 12 monthly forecast values for 2018 year are calculated by taking into account the Model II considered appropriate for the data. The forecasted values calculated, standard errors of these forecasted values and 95% confidence limits are given

---

[6] A criterion of model goodness which judges a model by how close its fitted values tend to be to the true values.

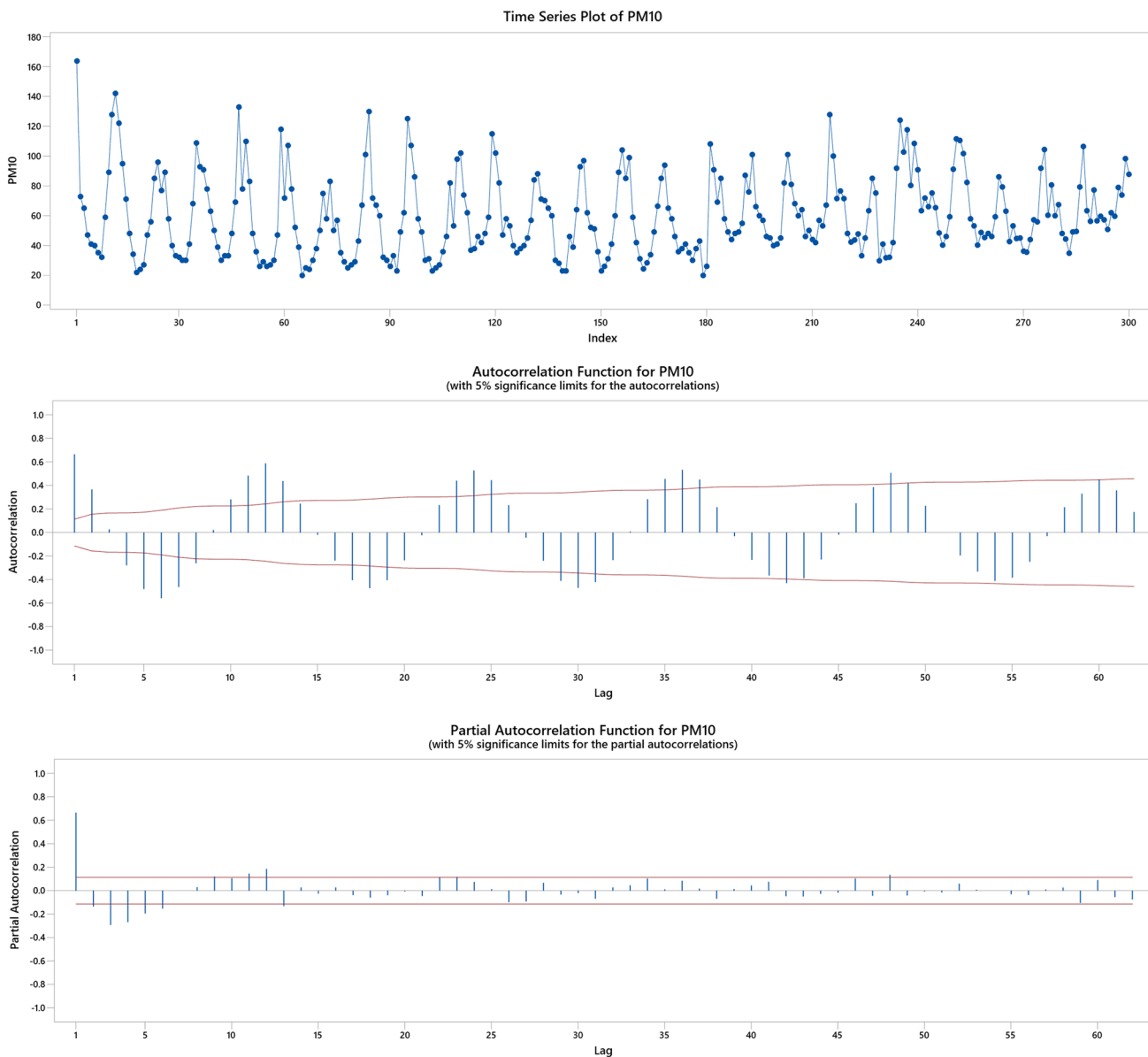[7] The unit root test which is proposed by Phillips and Perron (1988).

Fig. 1 Time series graphs of the PM_{10} data

in Table 7. When the forecast values are examined, it is expected that the highest value will be observed in November. It is followed by December and February.

Due to the advantages offered by the unit root test based on the periodograms, the stationarity of the series is also tested with periodograms. For Model II, the periodogram value and estimate of variance are calculated as $I_n(w_1) = 6133.66$ and $\hat{\sigma}_n^2 = 322.2444$, respectively. So, the value of the test statistic computed as

$$T_n(w_1) = 2(1 - \cos(w_1)) * I_n(w_1)/\hat{\sigma}_n^2 = 0.008349. \quad (19)$$

Following Akdi and Dickey (1998), the critical values of the test statistic are given as

$$P(T_n(w_1) \leq 0.0348) = 0.01, \quad P(T_n(w_1) \leq 0.178) = 0.05,$$
$$P(T_n(w_1) \leq 0.368) = 0.10. \quad (20)$$

So, the null hypothesis that the series is not stationary has been rejected at the significant levels $\alpha = 0.01$ and $\alpha = 0.05$ ($T_n(w_1) = 0.008349 < 0.178$ and $T_n(w_1) = 0.008349 < 0.0348$).

Table 1 Monthly average PM_{10} quantities

Month	N	Mean	Standard deviation
1	25	**85.2960000**	26.5522516
2	25	71.9480000	16.9131142
3	25	58.6680000	9.2753131
4	25	49.1680000	13.7989951
5	25	39.5440000	10.5722546
6	25	37.7360000	14.8041064
7	25	38.7480000	20.0870008
8	25	39.4440000	16.6427932
9	25	50.3400000	16.9200079
10	25	63.4280000	13.7862709
11	25	**96.6720000**	24.1641870
12	25	**88.1880000**	24.0925874

The bold expressions indicate the highest 3 PM_{10} averages by months for 1993–2018

Table 2 Generalized linear model procedure

Source	Degrees of freedom	Sum of squares	Mean squares	F value	$Pr > F$
Model	11	123362.2180	11214.7471	34.33	<0.0001
Error	288	94071.1712	326.6360		

Periodicity

We then investigate whether there is hidden periodicity in the series by means of periodograms. Figure 2 displays the periodogram against frequencies from 0 to π.

Figure 2 suggests that one of the periodograms differs significantly from the others. This can be regarded as a sign of a possible periodicity in the data (Wei 2006). The largest 5-periodogram values obtained from the data, the frequencies that corresponded to those and their periods are given in Table 8. Also, the total value of the periodograms in all frequencies is calculated as

The statistic V is

$$V = \left[\sum_{k=1}^{n/2} I_n(w_k) \right]^{-1} \max\{ I_n(w_k) \} = \frac{113026.59}{218213.92} = 0.51796228,$$

(21)

and the critical values that corresponded to the statistic V, with the approximate value $P(V > c_\alpha) = \alpha \cong m(1 - c_\alpha)^{m-1}$

Wei (2006), are calculated as $c_\alpha = 1 - (\alpha/m)^{1/(m-1)}$, where m is $n/2$ and is defined by

$$m = \begin{cases} (n-1)/2, & \text{if } n \text{ is odd} \\ (n/2) - 1, & \text{if } n \text{ is even.} \end{cases}$$

(22)

In order to look whether the other frequencies have the periodicity or not, the following formula is used:

$$V_i = \frac{I_n(w_{(i)})}{\sum_{k=1}^{n/2} I_n(w_k) - \sum_{k=1}^{i-1} I_n(w_{(k)})},$$

(23)

where $I_n(w_{(i)})$ for $i = 1, 2, 3, 4, 5$ are the periodogram values from big to small. Herefrom, $V_1 = 0.518$, $V_2 = 0.088$, $V_3 = 0.064$, $V_4 = 0.063$ and $V_5 = 0.043$ are obtained. For $\alpha = 0.01, \alpha = 0.05$ and $\alpha = 0.10$, the critical values are $c_{0.01} = 0.0624972682$, $c_{0.05} = 0.0523158530$ ve $c_{0.10} = 0.047896912$.

If $V > c_\alpha$, it will be rejected the null hypothesis that the dataset does not contain the periodic component and it will be decided that the series has a periodic component. According to the obtained results, since $V = 0.5180 > c_{0.01} = 0.0625$, it is concluded that there is the periodic component in the data. The period of the series is obtained as $P = 2\pi/w_k = 2\pi/0.52360 = 12$. This is an expected result because the data are monthly.

On the other hand, although it is $V_1 > c_\alpha, V_2 > c_\alpha, V_3 > c_\alpha$ and $V_4 > c_\alpha$, it is $V_5 < c_\alpha$. This shows that there is not a periodicity at the frequency corresponding to the fifth largest periodogram value. Therefore, the periodic components corresponding to the first four frequencies must be considered.

The harmonic regression model

Considering that the model is stationary and the data contain the periodic component, it is assumed that the following regression model (Trigonometric Model I) is appropriate for the data:

$$\begin{aligned} Y_t = {} & \mu + A_1 \cos\left(\frac{2\pi t}{12}\right) + B_1 \sin\left(\frac{2\pi t}{12}\right) \\ & + A_2 \cos\left(\frac{2\pi t}{12.5}\right) + B_2 \sin\left(\frac{2\pi t}{12.5}\right) \\ & + A_3 \cos\left(\frac{2\pi t}{300}\right) + B_3 \sin\left(\frac{2\pi t}{300}\right) + A_4 \cos\left(\frac{2\pi t}{6}\right) \\ & + B_4 \sin\left(\frac{2\pi t}{6}\right) + e_t, \quad t = 1, 2, \dots, 300. \end{aligned}$$

(24)

Estimates of model parameters are given in Table 9.

Table 3 Values of the AIC statistics

1	2	3	4	12	13	(4)	(12)	(1,12)	(1,12,13)
2654.46	2651.16	2625.84	2605.44	2577.06	**2573.59**	2806.71	2700.64	2596.64	2587.93
(404.91)	(399.16)	(365.65)	(340.49)	(301.86)	(297.44)	(672.62)	(472.29)	(332.83)	(322.24)

AIC values obtained using PROC ARIMA in SAS. The estimated variance for each model is given in brackets. The bold expression indicates the smallest AIC value among fitted models

Table 4 Parameter estimates for the Model I

Variable	Degrees of freedom	Parameter estimate	Standard error	t value	Pr > |t|
(a) The results of OLS (PROC REG)					
Intercept	1	28.99411	7.45407	3.89	**0.0001**
pm101	1	0.54602	0.05963	9.16	**<0.0001**
pm102	1	0.06291	0.06553	0.96	0.3379
pm103	1	−0.05727	0.06536	−0.88	0.3817
pm104	1	−0.08208	0.06549	−1.25	0.2112
pm105	1	−0.06870	0.06565	−1.05	0.2962
pm106	1	−0.09516	0.06569	−1.45	0.1486
pm107	1	0.02964	0.06587	0.45	0.6531
pm108	1	−0.03936	0.06566	−0.60	0.5494
pm109	1	0.05669	0.06558	0.86	0.3881
pm1010	1	−0.02932	0.06554	−0.45	0.6550
pm1011	1	0.04369	0.06563	0.67	0.5061
pm1012	1	0.30767	0.06491	4.74	**<0.0001**
pm1013	1	−0.16463	0.05686	−2.90	**0.0041**

Parameter	Estimate	Standard error	t value	Approx. Pr > |t|	Lag
(b) The results of PROC ARIMA					
MU	60.93544	1.99589	30.53	**<0.0001**	0
AR1, 1	0.52006	0.05863	8.87	**<0.0001**	1
AR1, 2	0.07608	0.06466	1.18	0.2403	2
AR1, 3	−0.06769	0.06473	−1.05	0.2965	3
AR1, 4	−0.08130	0.06491	−1.25	0.2114	4
AR1, 5	−0.05765	0.06504	−0.89	0.3762	5
AR1, 6	−0.09551	0.06515	−1.47	0.1438	6
AR1, 7	−0.00862	0.06540	−0.13	0.8952	7
AR1, 8	−0.02365	0.06518	−0.36	0.7170	8
AR1, 9	0.05203	0.06510	0.80	0.4249	9
AR1, 10	0.00652	0.06502	0.10	0.9203	10
AR1, 11	0.06211	0.06507	0.95	0.3406	11
AR1, 12	0.25495	0.06513	3.91	**0.0001**	12
AR1, 13	−0.13607	0.05918	−2.30	**0.0222**	13

The bold expressions indicate statistically significant parameters at 0.05 level of significance

Table 5 Parameter estimates for the Model II

Variable	Degrees of freedom	Parameter estimate	Standard error	t value	Pr > |t|
(a) The results of PROC REG					
Intercept	1	9.63705	2.73337	3.53	**0.0005**
pm101	1	0.57997	0.04876	11.90	**<0.0001**
pm1012	1	0.47482	0.05028	9.44	**<0.0001**
pm1013	1	−0.21874	0.05506	−3.97	**<0.0001**

Parameter	Estimate	Standard error	t value	Approx. Pr > |t|	Lag
(b) The results of PROC ARIMA					
MU	106.86832	12.31366	8.68	**<0.0001**	0
AR1, 1	0.66699	0.04583	14.55	**<0.0001**	1
AR1, 2	0.50664	0.05130	9.88	**<0.0001**	12
AR1, 3	−0.20195	0.05789	−3.49	**0.0006**	13

The bold expressions indicate statistically significant parameters at 0.05 level of significance

Table 6 Results of the ADF and PP unit root test

	t-statistics	Prob.
(a) The results of the ADF unit root test		
Augmented Dickey–Fuller test statistic	− 10.98959	0.0000
Test critical values:		
1% level	− 3.452442	
5% level	− 2.871161	
10% level	− 2.571968	
	Adj. t-Stat.	Prob.
(b) The results of the PP unit root test		
Phillips–Perron test statistic	− 6.808084	0.0000
Test critical values:		
1% level	− 3.452066	
5% level	− 2.870996	
10% level	− 2.571880	

Table 7 PM_{10} forecasts for 2018 year (PROC ARIMA)

Obs.	Forecast	Std. error	95% confidence limits	
The forecast values (2018:01–2018:12)				
301	77.1865	17.9512	42.0028	112.3701
302	**82.3933**	21.5779	40.1014	124.6851
303	70.9252	23.0083	25.8298	116.0206
304	69.1382	23.6168	22.8500	115.4263
305	66.0334	23.8826	19.2244	112.8424
306	61.3262	23.9999	14.2874	108.3651
307	65.0624	24.0518	17.9217	112.2032
308	64.1982	24.0749	17.0122	111.3842
309	73.7936	24.0852	26.5875	120.9997
310	73.7831	24.0898	26.5680	120.9982
311	**87.0973**	24.0918	39.8782	134.3163
312	**85.7507**	24.0927	38.5298	132.9715

The bold expressions indicate the highest 3 PM_{10} forecasts by months for 2018

Based on the p values in Table 9, it is decided that the appropriate model (Trigonometric Model II) is

$$Y_t = \mu + A_1 \cos\left(\frac{2\pi t}{12}\right) + B_1 \sin\left(\frac{2\pi t}{12}\right) + A_2 \cos\left(\frac{2\pi t}{12.5}\right)$$
$$+ B_2 \sin\left(\frac{2\pi t}{12.5}\right) + A_3 \cos\left(\frac{2\pi t}{300}\right) + B_3 \sin\left(\frac{2\pi t}{300}\right)$$
$$+ A_4 \cos\left(\frac{2\pi t}{6}\right) + e_t, \quad t = 1, 2, \dots, 300.$$

(25)

The parameter estimates obtained according to this model are as given in Table 10.

Based on the Trigonometric Model II, the following prediction model is established for the monthly PM_{10} data:

$$\hat{Y}_t = 59.93 + 27.27 \cos\left(\frac{2\pi t}{12}\right) + 3.14 \sin\left(\frac{2\pi t}{12}\right) + 3.71 \cos\left(\frac{2\pi t}{12.5}\right)$$
$$- 6.93 \sin\left(\frac{2\pi t}{12.5}\right) + 4.76 \cos\left(\frac{2\pi t}{300}\right) - 4.27 \sin\left(\frac{2\pi t}{300}\right)$$
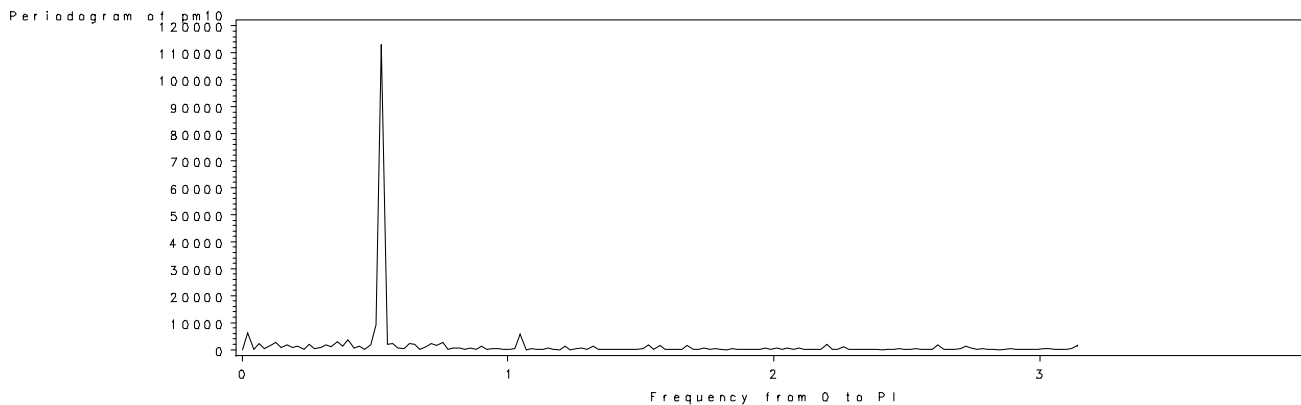$$+ 5.84 \cos\left(\frac{2\pi t}{6}\right).$$

(26)

Fig. 2 Graph of the periodograms against the frequencies

Table 8 Largest 5-periodogram values

k	Frequency (w_k)	$I_n(w_k)$	Period
26	0.52360	113026.59	12.000
25	0.50265	9273.12	12.500
2	0.02094	6133.66	300.000
51	1.04720	5678.46	6.000
20	0.39794	3636.59	15.789

The graph of the PM_{10} values observed and of the prediction values which were calculated according to this model is shown in Fig. 3a. Also, the graph of the PM_{10} values observed and of the prediction values which were calculated according to the Model II is shown in Fig. 3b.

The monthly average PM_{10} values observed for the last 23 years and the monthly average prediction values obtained from the ARIMA (Model II) and the harmonic regression (Trigonometric Model II) are presented in Fig. 3c. Note that the observations in the first 13 months are not taken into consideration in predictions, so the averages of the last 23 years are covered in the graphs.

As shown in Fig. 3c, the monthly average prediction values obtained by the harmonic regression are the closer to the real average values in comparison with the monthly average prediction values obtained by the ARIMA. The sum of the squares of the distances between these averages and the realized averages is calculated by

$$SS(\text{ARIMA}) = \sum_{i=1}^{12} (\bar{y}_i - \bar{y}_{a,i})^2 = 470.18105,$$

$$SS(\text{HARMONIC}) = \sum_{i=1}^{12} (\bar{y}_i - \bar{y}_{h,i})^2 = 279.63572,$$

(27)

where $\bar{y}_i$'s are the monthly realized averages, $\bar{y}_{a,i}$'s are the monthly average predictions obtained by the ARIMA and $\bar{y}_{h,i}$'s are the monthly average predictions obtained by the harmonic regression. Again, when the sums of squares are examined, it is concluded that the prediction values obtained by the harmonic regression are closer to the realized values in comparison with the prediction values obtained by ARIMA. This result is more apparent in Table 11, in which are present the realized values for the first two months of 2018 and the forecast values which were calculated by taking into account the prediction values obtained by both methods.

When Fig. 3a and b is examined, it is seen that the predicted values obtained for both models are very close to the observed values.

When the forecast values given above are examined, it is seen that the forecasts obtained from the both models are lower in the summer months (May, June, July and August). The result that the air pollution values observed in the summer months are lower than the other months is an expected result. The forecast values obtained from the harmonic model are much closer to the monthly average, whereas the forecast values obtained

Table 9 Parameter estimates for the Trigonometric Model I

Variable	Degrees of freedom	Parameter estimate	Standard error	t value	Pr > |t|
Intercept	1	59.93167	0.97695	61.35	**< 0.0001**
c1	1	27.27001	1.38161	19.74	**< 0.0001**
s1	1	3.13963	1.38161	2.27	**0.0238**
c2	1	3.71019	1.38161	2.69	**0.0077**
s2	1	− 6.93219	1.38161	− 5.02	**< 0.0001**
c3	1	4.76261	1.38161	3.45	**0.0007**
s3	1	− 4.26715	1.38161	− 3.09	**0.0022**
c4	1	5.84200	1.38161	4.23	**< 0.0001**
s4	1	− 1.93066	1.38161	− 1.40	0.1634

The bold expressions indicate statistically significant parameters at 0.05 level of significance

Table 10 Parameter estimates for the Trigonometric Model II

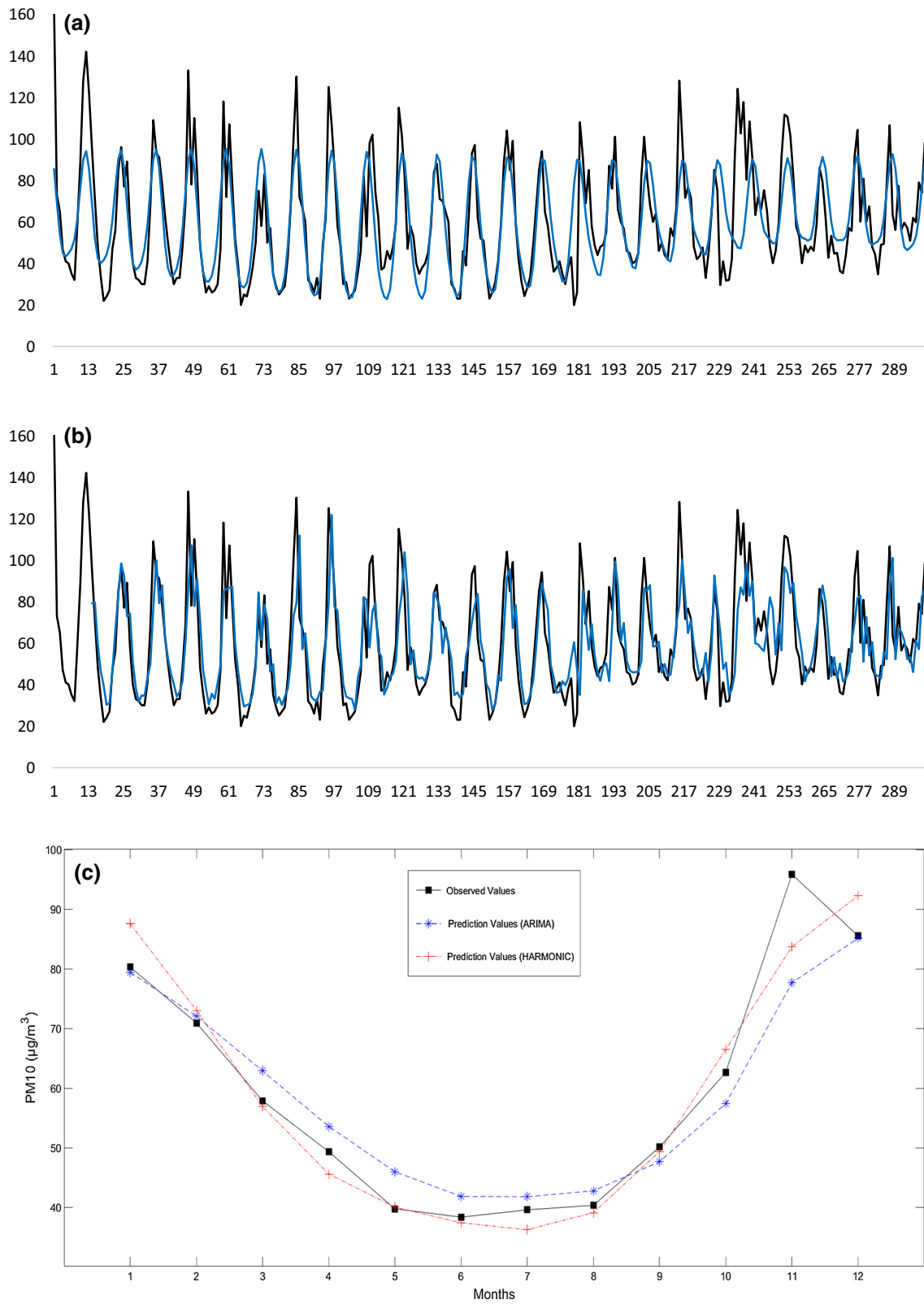| Variable	Degrees of freedom	Parameter estimate	Standard error	t value	Pr > |t|
Intercept	1	59.93167	0.97854	61.25	< 0.0001
c1	1	27.27001	1.38387	19.71	< 0.0001
s1	1	3.13963	1.38387	2.27	0.0240
c2	1	3.71019	1.38387	2.68	0.0078
s2	1	− 6.93219	1.38387	− 5.01	< 0.0001
c3	1	4.76261	1.38387	3.44	0.0007
s3	1	− 4.26715	1.38387	− 3.08	0.0022
c4	1	5.84200	1.38387	4.22	< 0.0001

Fig. 3 **a** Observed PM_{10} values and predictions of the Trigonometric Model II. **b** Observed PM_{10} values and predictions of the Model II. **c** The comparison of the monthly averages

Table 11 PM$_{10}$ values forecasted for the 2018 year

Month	Monthly average	Forecast (ARIMA)	Forecast (harmonic regression)	Realized value
January	85.296	77.1865	92.624	85.296
February	71.948	82.3933	74.073	72.060
March	58.668	70.9252	55.019	*
April	49.168	69.1382	42.625	*
May	39.544	66.0334	38.021	*
June	37.736	61.3262	38.148	*
July	38.748	65.0624	40.855	*
August	39.444	64.1982	47.607	*
September	50.340	73.7936	60.928	*
October	63.428	73.7831	79.426	*
November	96.672	87.0973	96.009	*
December	88.188	85.7507	101.915	*
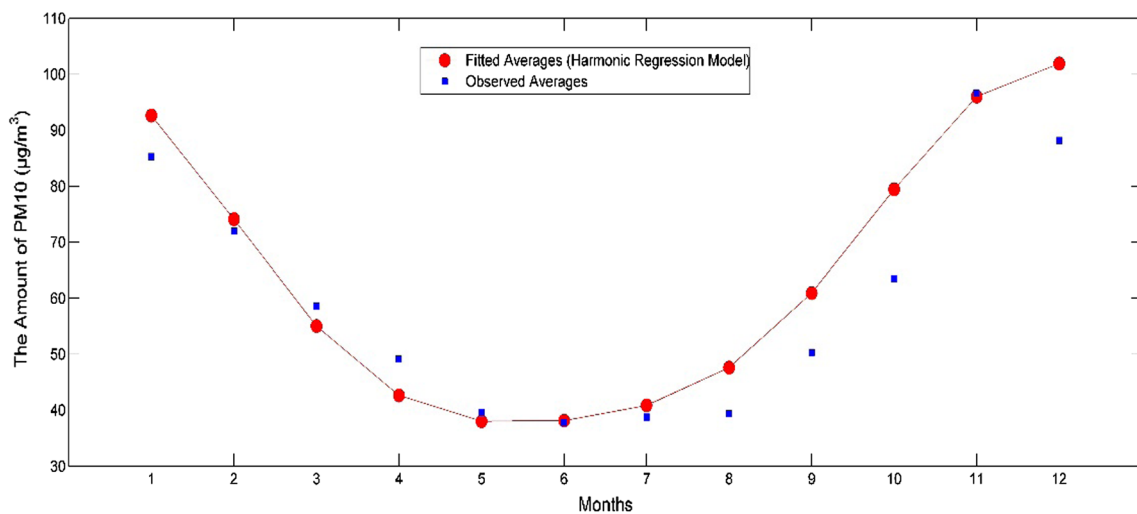
Fig. 4 Monthly averages and the graph of prediction equation (the harmonic regression model)

with ARIMA are well above the monthly average. Accordingly, it can be said that the harmonic regression model is more consistent and reliable than the ARIMA. In fact, this result can be also seen in Fig. 3c. The monthly average values and the graph of prediction equation are presented in Fig. 4.

Results and discussion

The model estimates and its yielding forecasts we presented in Sect. 4 were said to be superior to those produced by rivalling frameworks as evidenced by the comparison of alternative forecasts with actuals. Equivalently, the analytical approach of this paper comes with sizable benefits for the analysts and policy makers. It is salutary that these benefits are further augmented by the reduced costs of modeling, i.e., modeling as a professional effort. Among those, absence of the need for model assumptions seems to be an appealing one for any researcher

or analyst. The ability of one, especially of the field specialist, to compute periodograms without reference to model assumptions provides a certain improvement to reliability of projection practices of institutions. In a similar fashion, as opposed to the standard unit root tests that mandate the preliminary estimation of model parameters, the approach maintained here removes that need except for the variance of the white noise series.

The third advantage, from a purely academic rather than field work perspective, is the knowledge of the distribution of the statistic $T_n(w_k)$ under both H_0 and H_a, implying the existence of the analytical power function for the test. Independence of the critical values of the test statistic from the sample size, in addition, provides another strength.

Finally, the periodograms, being calculated through trigonometric transformations of data, can practically capture any periodic components, so promising more meaningful and accurate results. Over a practical domain of work, analysts or policy makers can easily obtain forecasts via our framework

simply by substituting a particular value of t, the period of forecast, into the estimated function.

In a nutshell, the harmonic regression approach as presented in this paper seems to be a source of ease and flexibility on the side of the field practitioners while providing the modelers with desired statistical properties.

With regard to the maintenance of a univariate modeling scheme, we again develop a benefit–cost perspective. Following the widespread principle of parsimony, also called the Occam's razor, among the models that yield similar explanatory or forecasting performances, the one(s) with lesser explanatory factors are to be preferred. This simple-looking yet strong principle ensures that a researcher could avoid redundant factors in a well-guided manner. In the current case as presented in this study, our pure time series approach based on periodograms departs from the knowledge of (1) embodiment of the periodic and non-periodic effects of all related factors in a time series of interest and (2) ability to express any time series as a sum of properly defined sinusoidal. As a matter of fact, these allowed us to produce quality forecasts even in the absence of explicit reference to potentially relevant factors like weather conditions or the behavior of other pollutants.

References

Akdi Y, Dickey DA (1998) Periodograms of unit root time series: distributions and tests. Commun Stat Theory Methods 27(1):69–87

Akdi Y, Dickey DA (1999) Periodograms of seasonal time series with a unit root. Istatistik J Turk Stat Assoc 2(3):153–162

Bayram H, Dikensoy Ö (2006) Hava Kirliliği ve Solunum Sağlığına Etkileri. Tüberküloz ve Toraks Dergisi 54(1):80–89

Berktaş BM, Bircan A (2003) Effects of atmospheric sulphur dioxide and particulate matter concentrations on emergency room admissions due to asthma in Ankara. Tüberküloz ve Toraks Dergisi 51(3):231–238

Bozkurt F, Altay ŞY, Yağanoğlu M (2015) Yapay Sinir Ağları ile Ankara İlinde Hava Kalitesi Sağlık İndeksi Tahmini. Natl Congr Manag Inf Syst 1(1):321–331

Brockwell PJ, Davis RA (1987) Time series: theory and methods. Springer, New York

Cavkaytar Ö, Soyer ÖU, Şekerel BE (2013) Türkiye'de Hava Kirliliğinden Kaynaklanan Sağlık Sorunları. Hava Kirliliği Araştırmaları Dergisi 2:105–111

Chamber of Environmental Engineers—Çevre Mühendisleri Odası (2018) Raporlar http://www.cmo.org.tr/odamiz/raporlar.php. Accessed 14 Feb 2018

Çiçek İ, Türkoğlu N, Gürgen G (2004) Ankara'da Hava Kirliliğinin İstatistiksel Analizi. Fırat Üniversitesi Sosyal Bilimler Dergisi 14(2):1–18

Dickey DA, Hasza DP, Fuller WA (1984) Testing for unit roots in seasonal time series. J Am Stat Assoc 79(386):355–367

European Commission—Environment (2018) Air quality—time extensions. http://ec.europa.eu/environment/air/quality/time_extensions.htm. Accessed 6 Jan 2018

Fuller WA (1996) Introduction to statistical time series. Wiley, New York

Ganguly R, Sharma D, Kumar P (2019) Trend analysis of observational PM10 concentrations in Shimla City, India. Sustain Cities Soc 51:101719

Genç DD, Yeşilyurt C, Tuncel G (2010) Air pollution forecasting in Ankara, Turkey using air pollution index and its relation to assimilative capacity of the atmosphere. Environ Monit Assess 166:11–27

Hylleberg S, Engle RF, Granger CWJ, Yoo BS (1990) Seasonal integration and cointegration. J Econom 44:215–238

Kaplan Y, Saray U, Azkeskin E (2014) Hava Kirliliğine Neden Olan PM10 ve SO2 Maddesinin Yapay Sinir Ağı Kullanılarak Tahmininin Yapılması ve Hata Oranının Hesaplanması. Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi 14:1–6

Koutrakis et al (2005) Analysis of PM10, PM2.5, and PM2.5–10 concentrations in Santiago, Chile, from 1989 to 2001. J Air Waste Manag Assoc 55:342–351

Kurt A, Gulbagci B, Karaca F, Alagha O (2008) An online air pollution forecasting system using neural networks. Environ Int 3:592–598

Li H, Shi X (2016) Data driven based PM2.5 concentration forecasting. Adv Biol Sci Res 3:301–304

Ostro BD, Hurley S, Lipsett MJ (1999) Air pollution and daily mortality in the CoachellaValley, California: a study of PM10 dominated by coarse particles. Environ Res Sect A 81:231–238

Phillips PC, Perron P (1988) Testing for a unit root in time series regression. Biometrika 75(2):335–346

Pope CA, Dockery DW (1992) Acute health effects of PM10 pollution on symptomatic and asymptomatic children. Am J Respir Crit Care Med 145(5):1123–1128

Said SE, Dickey DA (1985) Hypothesis testing in ARIMA(p,1, q) models. J Am Stat Assoc 80(390):369–374

Saral A (2000) Hava Kirliliğinin Yapay Sinir Ağları Yöntemiyle Modellenmesi ve Tahmini. Ph.D. Dissertation, Institute of Sciences, Yıldız Technical University

Silva LT (2015) Environmental quality health index for cities. Habitat Int 45:29–35

Turgut D, Temiz İ (2015) Time series analysis and forecasting for air pollution in Ankara: a Box-Jenkins approach. Alphanumeric J 3(2):131–138

TurkStat—Turkish Statistical Institute (2018) Address based population registration system results. https://biruni.tuik.gov.tr/medas/?kn=95&locale=en. Accessed 3 Feb 2018

Wei WWS (2006) Time series analysis: univariate and multivariate methods. Pearson Education, USA

WHO—World Health Organization (2018) Ambient (outdoor) air quality and health. http://www.who.int/mediacentre/factsheets/fs313/en/. Accessed 6 Jan 2018

Yüksek et al (2007) Sivas İlinde Yapay Sinir Ağları ile Hava Kalitesi Modelinin Oluşturulması Üzerine Bir Uygulama. C.Ü. İktisadi ve İdari Bilimler Dergisi 8(1):97–112