CSAWAC 43 (7) 967–1114 (2015) · Vol. 43 · No. 7 · July 2015

CLEAN
Soil Air Water

Renewables

Sustainability

Environmental Monitoring

7|2015

www.clean-journal.com

WILEY

CLEAN
Soil Air Water

Fatih Taşpınar

Environmental Engineering Department, Düzce University, Düzce, Turkey

Research Article

Time Series Models for Air Pollution Modelling Considering the Shift to Natural Gas in a Turkish City

Fossil fuel utilization for residential heating is still a major source of fine particulate matter (PM_{10}) and sulphur dioxide (SO_2), despite the increasing consumption of natural gas in some cities in Turkey. In the present study, PM_{10} and SO_2 air pollution and residential natural gas consumption (RNGC) were modelled by various multi-parameter time series modelling methods (TSMs). To estimate short-term pollution levels considering the future estimates of RNGC and meteorological factors, a time series dataset was designed for the years 2007–2013. Factor analysis was also performed to aid in the selection of variables for constructing TSMs. The error measures and coefficient of determination (R^2) were used to evaluate forecasting accuracy of the models constructed. In the short-term estimation of RNGC, PM_{10}, and SO_2 for 2014–2015, temperature dependent ARIMAX(1,1,2) ($R^2 = 0.944$) and RNGC and meteorological factors dependent SARIMAX$(0,1,1)(1,1,0)_{12}$ ($R^2 = 0.761$) and ARIMAX(1,1,0) ($R^2 = 0.698$) models, respectively, yielded the best fitting scores and accuracy measures. The models performed well in reflecting the time series data and thus, could be utilized in energy planning for sustainable development concerning environmental decision making and short-term air quality forecasting for public health.

Keywords: Air quality forecasting; Energy consumption; Factor analysis; PM_{10}; SO_2

Received: June 19, 2014; *revised:* September 21, 2014; *accepted:* November 7, 2014

DOI: 10.1002/clen.201400461

1 Introduction

For decades, an increase in energy consumption due to rapid industrialization and population growth has led to the deterioration of ambient air quality all over the world. Since the second half of the 20th century, environmental concerns regarding air pollution have encouraged wider access to natural gas (NG) among other primary energy sources. On a global scale, adverse effects on human health and air quality may be seen as a consequence of anthropogenic activities responsible for economic development and energy dependency [1, 2]. Mobile and stationary fossil fuel combustion is well known as a major source of anthropogenic gaseous pollutants and particulate matter. There are numerous studies in the field of air pollution modelling and assessment concerning particles with an aerodynamic diameter $\leq 10\,\mu m$, PM_{10}, and SO_2. Both are mostly combustion-originated and among the six criteria pollutants according to the National Ambient Air Quality Standards due to their adverse effects on human health [3–6].

Air pollution phenomena are mostly investigated through time series data that includes some degree of randomness due to atmospheric events during varied time periods. Many studies have revealed that the air quality data are stochastic time series by making short-term estimations possible by exploring historical data patterns [7–9]. The most widely employed methods are the non-seasonal and seasonal autoregressive integrated moving average (e.g., ARIMA, SARIMA) models in time series analysis, which are based on a single target variable, namely univariate models [10–12]. In the case of conventional air pollutants (APs) such as PM_{10}, CO, SO_2, and O_3, seasonal time series models have been successfully applied to monitor datasets which are daily or monthly based [13–15]. The resulting model quality generally relies on individual experience, knowledge of time series analysis methods in the model identification stage and visualization of time series forecasting plots, which leads to establish several models for the same dataset.

In some Turkish cities, air pollution due to residential heating is still included among the listed pollution sources and has become a great environmental concern. However, since 2000, the extensive use of NG as a substitute for coal and lignite for residential and industrial heating purposes has considerably improved the overall air quality of Turkey [16–18]. Thus, many measures have been taken to improve air quality, two examples being the encouragement of energy alternatives by supporting investment in renewable energy

Correspondence: Dr. F. Taşpınar, Environmental Engineering Department, Düzce University, Düzce 81620, Turkey
E-mail: fatihtaspinar@duzce.edu.tr

Abbreviations: **AP**, air pollutant; **APC**, concentration of air pollutants; **AT**, air temperature; **DES**, double exponential smoothing; **FA**, factor analysis; **NBIC**, normalized Bayesian information criterion; **NG**, natural gas; **PM_{10}**, particulate matter up to 10 μm in diameter; **RH**, relative humidity; **RMSE**, root mean squared error; **RNGC**, residential natural consumption; **(S)ARIMAX**, (seasonal) autoregressive integrated moving average with exogenous variables; **SES**, single exponential smoothing; **TES**, triple exponential smoothing; **TSM**, time series method; **WS**, wind speed sources and proposing regulations for substituting NG for other solid fossil or biomass fuels [19]. Energy utilization and its environmental impacts, especially in large Turkish cities, have been analyzed and extensively studied [20–22]. However, the cities like Düzce were characterized by semi-urban features and have not been sufficiently examined in terms of fuel shifting.

A great number of researchers employed univariate time series models with a satisfactory model performance. Although these models are able to simulate historical data well, they do not consider any external inputs that may greatly contribute to the explainable variance on the model variable. Hence, in order to extend the abilities of univariate time series modelling approaches by involving exogenous variables in the models as input variable, particularly meteorological factors, their multivariate forms, e.g., ARIMAX and SARIMAX, have also been successfully applied to air pollution datasets for a given time period [23, 24]. However, the models given in the literature that were constructed in the forms of univariate or multivariate have been employed to predict the air pollutants levels within the limited time period of the dataset used. In other words, they have not been applied to estimate possible future data points after the actual data period. In addition, in case of particulate matter and sulphur dioxide pollution originating from solid or biomass fuel burning, any model was not established considering meteorological factors along with environmental benefits of prolonged NG consumption during heating season on air quality.

There are two primary objectives of this study: (1) the short-term estimation of local residential natural consumption (RNGC) by using time series methods (TSMs) based on meteorological factors, and (2) the prediction of future levels of PM_{10} and SO_2 depending on the RNGC estimates implementing multivariate TSMs for the province of Düzce. To implement the models and perform statistical analysis, we have used the functions in RStudio with forecasting package and RExcel (R3.1.0) (www.r-project.org/) and Mathworks© Matlab. Factor analysis (FA) was applied before modelling to reveal the hidden correlations among RNGC, meteorological variables and APs by principal factors that also assist in the selection of the model variables. A time series data set was designed covering the concentrations of PM_{10} and SO_2, RNGC, meteorological factors, and some socioeconomic indicators for 2007–2013. In the modelling stage, TSMs including ARIMAX and SARIMAX, smoothing methods and multiple regression were examined to produce better estimations for future levels of RNGC and APs for 2014–2015, and estimated short-term concentration of air pollutants (APCs) were interpreted considering shifting to NG for domestic heating on temporal air quality.

2 Materials and methods

2.1 Area of interest

The province of Düzce is located in north-western Turkey and is the eighty-first and the most recently created province of Turkey (Fig. 1). It is generally characterized by hilly to mountainous topography which covers about 85% of the province. Although the area is extensive, it has a relatively small population of about 346 500, according to the 2012 census (e.g. www.turkstat.gov.tr/) [25, 26]. The population is approximately evenly divided between rural and urban areas, but rural features are still dominant in the province. Because of the rough and hilly geography, the prevailing wind flow to the central district of Düzce is blocked, resulting in exposure to elevated levels of air pollutants. Due to temperature inversion, the weather in Düzce is foggy for about six months during winter time, from October to March. A low inversion layer above the downtown area of the city can be observed in winter and summer. This inversion leads to intensifying the air pollution, such as smog trapped close to the ground. Therefore, the pollution episodes are commonly associated with smog caused by the use of fossil fuels for heating and amplified by the inversion.

The November 1999 Düzce earthquake resulted in widespread destruction. After this disaster, the reurbanization and reconstruction process was accelerated by the state. In Düzce, residential heating is mostly provided by NG-based heating systems and stoves. Stoves which can burn solid fossil fuels such as coal, lignite, and biomass fuels like firewood and nutshells are used for space heating in 40–50% of households. Natural gas-based heating systems, such as NG combination-boilers or district heating, are used in 60–70% of

Figure 1. Location map of Düzce Province in Turkey.

CLEAN Soil Air Water

the homes, most of which were built after the earthquake. The use of NG-based heating systems in the residential sector rapidly increased after the installation of a pipeline system for gas transmission and distribution that was made operational in 2006.

2.2 Data used and explanatory statistics

A daily dataset for Düzce was composed of information about local RNGC, meteorological parameters, and APCs for 2007–2013. The RNGC data were obtained from the AKSA DERGAZ Gas Distribution Company [27] and the meteorological data from the General Directorate of Meteorological Affairs of Turkey. The APCs were taken from the Ministry of Environment and Urban Planning, using the online web service of the National Air Quality Monitoring Network of Turkey. The variables in the dataset consisted of RNGC (sm^3), air temperature (AT, °C), wind speed (WS, m/s), relative humidity (RH, %), and concentration levels of PM_{10} and SO_2 in $\mu g/m^3$. Table 1 summarizes the descriptive statistics of the variables and Fig. 2 illustrates their time series plots for 2007–2013.

Since mid-2006, a rising trend in the utilization of NG, peaking during the heating seasons, can be observed. The monthly average RNGC increased approximately five-fold from 1.18E6 sm^3 in 2007 to 10.52E6 sm^3 in 2012 over the past six years. Parallel to this, the PM_{10} and SO_2 levels exhibited a declining trend, as shown in Fig. 2. However, the concentration levels of SO_2 and PM_{10} still peak with the extended use of solid fossil fuels for heating during the cold months. The long-term daily mean concentration levels of PM_{10} and SO_2 were 93.8 ± 47.2 and 7.8 ± 4.5 $\mu g/m^3$, respectively.

Other fuels such as coal, firewood, and nutshells are used for residential heating in Düzce besides NG. Since mid-2008, NG and coal prices have risen by about 40%. Over the past six years, the general increase in fuel prices on average has been approximately 150% [28, 29]. We have no statistically proven official data reporting the exact number of households which utilize biomass and solid fossil fuels in Düzce. However, the number of registered residential customers of the local gas distribution company, AKSA, is on record, and this corresponds to about 60–70% of the households in the province. Hence, it can be assumed that the rest of the population, approximately 35 000–40 000 households, still utilize solid fossil and biomass fuels for residential heating. Nevertheless, there may also be some registered NG consumers of the company who do not use NG in their homes due to the high monthly cost of NG during the heating season. A recent fuel-consumption survey revealed that the main contributor to atmospheric pollution was the burning of low-quality fossil fuels for heating considering preferred fuels in previous years.

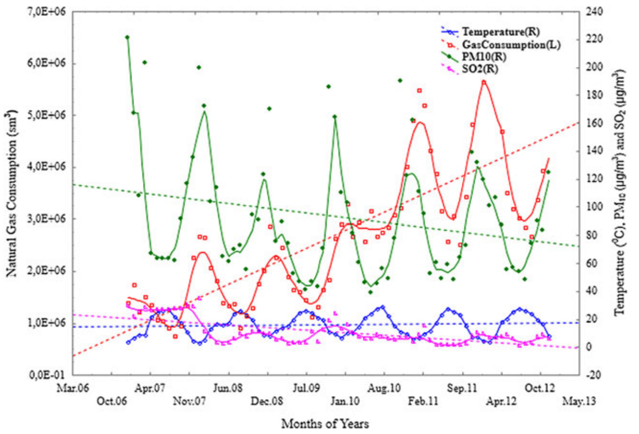

Figure 2. Temporal changes in the levels of monthly average RNGC, AT, and concentrations of PM_{10} and SO_2 for 2007–2013.

2.3 Factor analysis as an exploratory tool

The basic objective for FA implemented in this study was to determine the independent relationships between the variables in the dataset in terms of seasonal air pollution and to use factor groups to reveal the variables affecting RNGC and APs for modelling [30, 31]. FA was applied to the correlation coefficients of the variables by employing a principal component analysis with varimax rotation as the extraction method of factors. The variables with a communality value $\geq$0.5 were included in the final run to extract the principal factors. The factors with Eigenvalues $\geq$1.0 were retained for further analysis, as the Kaiser criterion suggests. Later, rotated factors with loading scores $\geq$0.7 were interpreted. By accounting for meteorological factors and RNGC, as well as APs, these factors were used to test the hypothesis assuming the source of pollutants is related to the extended use of solid fossil fuels for heating.

2.4 Predictions by time series modelling methods

TSMs in forecasting studies are used to predict future data points by revealing historical data pattern in the series. Widely used TSMs are ARIMA, SARIMA, and their multivariate forms ARIMAX and SARIMAX, and exponential smoothing models [32–35]. To obtain TSMs for Düzce in the estimation of RNGC and APCs, meteorological factors were involved in the models as well. Based on these variables, many models including ARIMAX(p,d, q), and its seasonal form SARIMAX(p, d, q)(P, D, Q) s, were examined. The non-negative integer elements p, d, and q used in the non-seasonal models refer to the order of autoregressive part (AR(p)) and the order of differencing (I(d))

Table 1. Descriptive statistics of the dataset (2007–2013)

Parameter	Maximum	Minimum	Range	Mean	SE	SD	Variance
Monthly average RNGC (sm^3)	10.52E6	1.18E6	9.32E6	4.3E6	269 986	2 290 903	5.25E+12
PM_{10} ($\mu g/m^3$)[a]	995	9	986	93.8	7.87	47.2	2230.95
SO_2 ($\mu g/m^3$)[a]	348	1	347	7.8	0.75	4.5	20.05
T (°C)[a]	42	−15	26	16.3	1.25	7.5	56.39
WS (m/s)[a]	21.8	0.1	21.7	1.3	0.11	0.7	0.46
RH (%)[a]	139.3	3.4	135.9	78.3	9.7	23.3	542.89

SE, standard error; SD, standard deviation.
[a] Daily based maximum and minimum values.

and moving average (MA(q)) parts of the models, respectively. Likewise, P, D, Q, and s in the seasonal models refer to the seasonal order of the autoregressive (SAR(P)), differencing ($I_S(D)$) and the moving average (SMA(Q)) parts, and the seasonal period of the model, respectively. Furthermore, exponential smoothing models were constructed to fit test models by investigating trends and seasonality in the time series data. The ARIMAX and SARIMAX models can be represented by the following equations, respectively:

$$y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \ldots + \beta_k X_{k,t} + \frac{(1-\theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)}{(1-\varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_p B^p)}\epsilon_t \quad (1)$$

$$y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \ldots + \beta_k X_{k,t} + \frac{(1-\theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)(1-\Theta_1 B^s - \Theta_2 B^{2s} - \ldots - \Theta_Q B^{Qs})}{(1-\varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_p B^p)(1-\Phi_1 B^s - \Phi_2 B^{2s} - \ldots - \Phi_P B^{Ps})}\epsilon_t \quad (2)$$

where y_t is the t-th observation of the dependent variable, $X_{1,t}, X_{2,t}, \ldots, X_{k,t}$ are the corresponding observations of the explanatory variables, β_0 is a constant, β_1, β_2, .. , β_k are the parameters of the regression part, and B is the backshift operator ($By_t = y_{t-1}$, $B^2 y_t = y_{t-2}$), ϵ_t is error residuals ($\sim$N(0,σ^2)), $\emptyset_1$, $\emptyset_2$, ..., $\emptyset_p$, Φ_1, Φ_2, and ..., Φ_p, θ_1, θ_2,..., θ_q and θ_1, θ_2, ..., θ_Q are the weights for the non-seasonal and seasonal, autoregressive terms and moving average terms, respectively. To test the lack of fit of TSMs, the Ljung–Box test was used as a model diagnostic tool and the most suitable model was selected according to normalized Bayesian information criteria (NBIC) [33–38].

The smoothing is another method used in time series analysis and is usually performed to investigate underlying patterns in the data, such as trends, seasonality, or both. In order to reduce irregularities in the time series data, smoothing techniques are used in the forecasting. The most frequently applied smoothing methods include the moving average, and single, double (Holt's method) and triple (Holt–Winter's method) exponential smoothing, so-called SES, DES, and TES, respectively [32, 36]. General forms of SES, DES, and TES, and forecasting equations are given as follows, respectively:

$$S_t = ay_t + (1-a)S_{t-1} \quad (3)$$

$$S_t = ay_t + (1-a)(S_{t-1} + b_{t-1}), b_t = \gamma(S_t + S_{t-1}) + (1-\gamma)b_{t-1} \quad (4)$$

$$S_t = a\frac{y_t}{I_{t-L}} + (1-a)(S_{t-1} + b_{t-1}), b_t = \gamma(S_t + S_{t-1}) + (1-\gamma)b_{t-1}, I_t = \beta\frac{y_t}{S_t} + (1-\beta)I_{t-L} \quad (5)$$

$$F_{t+m} = (S_t + mb_t)I_{t-L+m} \quad (6)$$

where y_t is the observation, S is the smoothed observation, b is the trend factor, I is the seasonal index, F is the forecast at m periods ahead, t is the time period index, and $0 \le \alpha, \beta, \gamma \le 1$ are the constants that must be estimated.

To obtain a stationary time series, differencing and logarithmic transformation were applied to dataset. Then, (S)ARIMAX and smoothing methods were used to construct a fitting model for short-term RNGC estimates on the basis of meteorological factors. Later, the best-fitting model in the prediction of short-term APCs depending on RNGC estimates and meteorological factors was identified. Multiple ordinary least squares regression was also implemented to obtain linear models of RNGC and APCs as benchmarks. The model with the lowest NBIC and the greatest non-significant Ljung–Box statistics ($p > 0.05$) was selected as the best among the others identified. The error measures root mean squared error (RMSE), mean absolute percentage error, and coefficient of determination (R^2) were also used to evaluate the forecasting accuracy of the models. Except for R^2, the lower values for these measures indicated a better-fitting model yielding the minimum forecasting error.

3 Results and discussion

3.1 The results of FA and the relationships between variables

The correlation coefficient values (r) used in FA are given in Table 2. The AT was correlated with a highly negative PM_{10} ($r = -0.83$) and with a moderately strong SO_2 ($r = -0.57$) and RNGC ($r = -0.68$), indicating the seasonal effect due to residential heating controlled by meteorological factors. The positive correlation of $r = 0.46$ between RNGC and PM_{10} may be explained by considering the extremely high increase in their levels during winter comparing to their annual averages, which also indicates a predominant feature of seasonal air pollution. The FA method over the correlation results was deemed acceptable, as the r values were generally >0.3.

However, the variable RH was removed from the final FA operation because its communality value was <0.5. Therefore, the five parameters, RNGC, SO_2, PM_{10}, AT, and WS were retained in the analysis. The FA resulted in extracting two factors which was numerically close to each other with the Eigenvalues ≥ 1. These factors with Eigenvalues of 2.12 and 1.99, accounted for 82.36% of the total variance. Table 3 shows the factor loadings with the communalities of the variables. The variables with a factor loading score ≥ 0.7 were left to evaluate the extracted factors. Bartlett's test of sphericity with a p-value <0.001 indicated the suitability of the data for latent factor detection.

Table 2. The cross correlation matrix of the variables in the dataset

Variable	RNGC	PM_{10}	SO_2	AT	WS	RH
RNGC	1.00	0.46	−0.31	−0.68	−0.67	−0.01
PM_{10}		1.00	0.53	−0.83	−0.43	0.42
SO_2			1.00	−0.57	−0.29	0.31
AT				1.00	0.18	−0.42
WS					1.00	−0.19
RH						1.00

Table 3. The results of FA after varimax rotation

Variable/Parameter	Extracted factor		Communality (>0.5)
	Factor-1	Factor-2	
RNGC	0.37	−0.93	0.87
PM_{10}	0.88	−0.28	0.86
SO_2	0.76	0.36	0.72
AT	−0.77	0.72	0.89
WS	0.39	0.78	0.77
Eigenvalues	2.12	1.99	
Explained variance (82.36%)	42.03%	40.33%	

The first factor, Factor-1, best explaining 42.03% of the total variance, was comprised of AT, PM_{10}, and SO_2. The change in the air temperature showed a fairly strong negative association, particularly with PM_{10}. Seasonal variations of APCs due to heating seasons with respect to non-heating seasons were indicated. Factor-1 represents the seasonality of air pollution governed by the seasonal temperature changes, explaining the reason for negative temperature loading on this factor with the others. Factor-1, considering the group of variables loaded on this factor, can thus be called as "seasonal pollution effect".

The second factor, Factor-2, explains about 40.33% of the total variations, grouping RNGC, WS, and AT. The most effective variable in this factor was RNGC, which showed a strong negative correlation with WS and AT. The correlation and factor loading score between RNGC and WS was negative because of the lesser WS levels observed during winter in Düzce comparing to its annual average and extremely high increase in RNGC, which is in contrast to summer. The extremely high decrease in natural gas consumption for domestic heating during warm seasons, and vice versa, of course resulted in a negative correlation between AT and RNGC. Therefore, taking into account the negative correlations between RNGC and AT, as well as WS, Factor-2 can be interpreted as "meteorological effect".

These factors indicating winter and summer periods, and the great difference in APCs obtained during heating and non-heating seasons revealed the seasonal air pollution. As expected, the interpretation of the extracted factors showed that the air pollution in Düzce is primarily due to heating during winter.

3.2 Temporal evaluations of average PM_{10} and SO_2 levels over RNGC

The substitution of NG was visualized as a means of achieving the abatement of air pollution. Temporal variations in PM_{10} and SO_2 levels against RNGC for the six-month periods of winter (November–April) and summer (May–October) are plotted in Fig. 3a and b for 2013–2015. Overall, the average monthly RNGC for the winter and summer periods increased from 2.05E6–6.83E6 sm^3, and the average values of RNGC were found to be approximately 5.56 and 3.37 sm^3, respectively. It can be seen that a reasonable decrease in the average levels of PM_{10} and SO_2 was established with increasing RNGC; however, peak emission levels can still occur, especially during cold months (Fig. 3a) due to extended solid fuel use for space heating. Hourly concentration levels provide more information about the seasonal pollution levels in the investigation. The average hourly concentrations of PM_{10} and SO_2 in winter were 134 ± 80 and $12 \pm 10\,\mu g/m^3$, whereas the summer hourly averages were 61 ± 36 and $6 \pm 5\,\mu g/m^3$, respectively. Average levels and exceedances showed the major pollutant species for the region to be particulates, which are directly connected to fuel consumption and local meteorology. Therefore, the estimation of short-term levels of APs may provide important information to authorities and researchers concerning local energy consumption and its effects on local air quality level.

3.3 Short-term estimates of RNGC and APCs using time series modelling

To estimate short-term RNGC, various (S)ARIMAX and exponential smoothing models were examined, considering model significance, error scores, and accuracy measures. The logarithms of RNGC (logRNGC) were used in these models in terms of stationary series. The variables AT, RH, and WS, were placed as exogenous variables in TSMs in the tests; however, only the variable AT was found to be significant ($p < 0.05$) in parameter estimates. The obtained models for short-term RNGC estimates and error measures are given in Table 4. In the construction of models, the order of the model is selected by plotting the ACFs and PACFs for model variable. To be able to input AT as an exogenous variable to the developed models for 2014–2015, a SARIMA$(0,0,1)(0,1,0)_{12}$ model ($R^2 = 0.811$, RMSE $= 3.266$, NBIC $= 2.378$) was fitted before modelling RNGC. This AT model is a seasonal model including MA(1) and $I_{12}(1)$ processes due to typical cycling patterns of air temperature. For RNGC modelling,

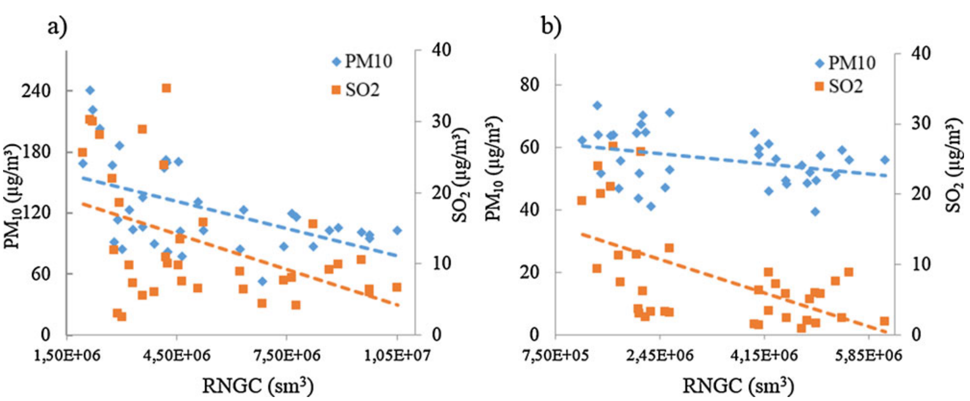

Figure 3. Temporal variations in average levels of APs with RNGC for 2007–2013: (a) Winter, (b) summer months average levels.

Table 4. The model statistics of TSMs in predicting short-term RNGC

Model	NBIC	Ljung-Box Q(18) with sig.	RMSE	MAPE	R^2
TES	−5.595	Q = 10.79, sig. = 0.89	0.57	0.68	0.944
ARIMA(1,1,2)	−4.657	Q = 36.36, sig. = 0.56	0.86	1.05	0.862
ARIMAX(1,1,1)	−5.218	Q = 10.90, sig. = 0.81	0.63	0.72	0.902
ARIMAX(1,1,2)	-5.171[a]	Q = 10.52, sig. = 0.77	0.62	0.72	0.917
ARIMAX(1,1,2)	−4.790	Q = 11.42, sig. = 0.73	0.81	0.84	0.842
SARIMAX(1,1,1)(1,1,1)$_{12}$	−2.532	Q = 11.74, sig. = 0.65	0.79	0.91	0.875
SARIMAX(1,1,2)(1,1,1)$_{12}$	−4.280	Q = 11.08, sig. = 0.61	0.80	0.90	0.878
SARIMAX(1,1,2)(1,1,0)$_{12}$	−4.481	Q = 11.88, sig. = 0.58	0.81	0.90	0.849
OLS regression [$AT_{(t-1)}$, $logRNGC_{(t-1)}$]	–	–	0.89	0.96	0.681

MAPE, mean absolute percentage error.
[a] The best score is bold

an order of 1 for the (S)AR process, i.e., AR(1), and an order of 1–2 for (S)MA process with differencing were identified and tested. The best scores with significant model parameters for the RNGC estimates were obtained by the TES ($R^2 = 0.944$, RMSE = 0.57, NBIC = −5.595) model and the AT dependent ARIMAX(1,1,2) model ($R^2 = 0.915$, RMSE = 0.62, NBIC = −5.171), which is based on AR(1) and MA(2) with one differencing process with the coefficients of $\emptyset_1$: −0.816, θ_1: −0.452 and θ_2: 0.267 for RNGC, respectively, and the AR(1) process with the coefficient of $\emptyset_1$: −0.273 for AT. The seasonal models produced bigger NBIC values and all the models tested required a differencing order of 1 to eliminate remaining trend and cycles.

In the case of RNGC, the NBIC value was the lowest in ARIMAX(1,1,2) model comparing to seasonal models and so, this model was selected for the short-term RNGC estimates. Based on the parameter estimates at significance level $p < 0.05$, the expanded forms of the AT and RNGC models can be written as follows:

$$AT_t = 0.253 + AT_{t-12} + (1 + 0.506)\epsilon_t \tag{7}$$

$$\begin{aligned} log(RNGC_t) = 0.617 + log(RNGC_{t-1}) \\ - 0.816(log(RNGC_{t-1} + log(RNGC_{t-2})) - AT_t - AT_{t-1} \\ + 0.273(AT_{t-1} - AT_{t-2}) + (1 + 0.452B - 0.267B^2)\epsilon_t \end{aligned} \tag{8}$$

Figure 4 displays the results of the short-term RNGC estimates by the ARIMAX(1,1,2) model along with the upper and lower confidence limits ($\alpha = 5\%$) and the estimates of AT by the SARIMA(0,0,1)(0,1,0)$_{12}$ model for 2013–2015. The predicted RNGC values appeared to follow the upturn and downturn of the observed series corresponding to the seasonal patterns of AT reasonably well. Over the study period, the minimum and maximum monthly RNGC estimates of the model were in the range of 5.81E6–20.40E6 sm^3, with a three-year monthly average of 11.29 sm^3. The estimated winter monthly RNGC averages for 2013–2015 were 9.32E6, 12.11E6, and 15.72E6 sm^3, which denotes the increasing trend of RNGC fitting with the past data pattern. Previous studies [34] showed that autoregressive models mostly have been applied as base approach and yielded reasonable scores in energy demand modelling with additional inputs.

Likewise, in the estimation of possible future levels of APCs based on RNGC and AT estimates, (S)ARIMAX and exponential smoothing models were examined. The models with significant scores for PM_{10} and SO_2 were identified and are shown in Table 5, along with error measures and model diagnostics.

In the short-term estimation of PM_{10} and SO_2 levels, RNGC and AT dependent models in the forms of SARIMAX(0,1,1)(1,1,0)$_{12}$ (RMSE = 21.38, $R^2 = 0.761$, NBIC = 6.495) and ARIMAX(1,1,0) (RMSE = 4.39, $R^2 = 0.698$, NBIC = 3.378), respectively, yielded the best scores among the other models tested. A seasonal model obtained for PM_{10} forecasting, with a MA(1) ($\emptyset_1$: −0.208) and SAR(1) (Φ_1: 0.098) process and non-seasonal and seasonal differencing order of 1, revealed the historical pattern of winter episodes over the years. However, the lower annual variations in average SO_2 levels comparing to PM_{10} resulted in selecting a non-seasonal time series model for it with an AR(1) ($\emptyset_1$: −0.129) process and a differencing order of 1. Based on the

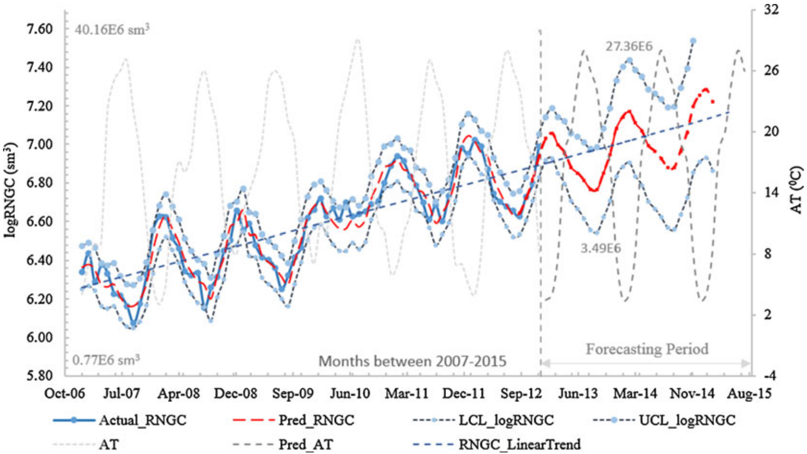

Figure 4. RNGC estimates of ARIMAX(1,1,2) model and variations in AT for 2013–2015.

CLEAN Soil Air Water

Table 5. TSMs for short-term PM_{10} and SO_2 predictions based on RNGC and AT estimates

	NBIC	Ljung-Box Q(18) and sig.	RMSE	MAPE	R^2
	Model for short-term PM_{10} estimates				
Model (dependents)					
DES	6.590	Q = 11.40, sig. = 0.784	25.42	19.19	0.656
$ARIMA(0,0,1)(1,0,0)_{12}$	7.727	Q = 17.80, sig. = 0.336	24.39	20.60	0.611
$SARIMAX(0,1,1)(1,0,0)_{12}$ [AT, logRNGC]	7.323	Q = 21.06, sig. = 0.167	32.45	26.69	0.444
$SARIMAX(0,0,1)(1,0,1)_{12}$ [logRNGC]	7.468	Q = 17.72, sig. = 0.274	33.57	25.13	0.433
$SARIMAX(0,1,1)(1,1,0)_{12}$ [AT, logRNGC]	**6.495**[a]	Q = 10.82, sig. = 0.896	21.38	16.19	0.761
$SARIMAX(0,0,1)(1,0,0)_{12}$ [logRNGC]	7.226	Q = 11.82, sig. = 0.756	28.12	19.52	0.547
$SARIMAX(0,0,1)(1,0,0)_{12}$ [AT, logRNGC]	7.203	Q = 11.83, sig. = 0.755	25.06	18.209	0.661
PM10_OLS_Regression [$AT_{(t-1)}$, $logRNGC_{(t-1)}$, $PM_{10(t-1)}$]	–	–	28.34.	23.78	0.651
Model for short-term SO_2 estimates					
DES	3.471	Q = 11.32, sig. = 0.716	4.54	44.75	0.627
ARIMAX(0,1,1) [AT, logRNGC]	4.676	Q = 14.53, sig. = 0.574	6.31	73.87	0.531
ARIMAX(1,1,0) [AT, logRNGC]	**3.378**[a]	Q = 10.83, sig. = 0.793	4.39	42.18	0.698
$SARIMA(1,1,1)(1,0,1)_{12}$ [AT, logRNGC]	4.872	Q = 14.408, sig. = 0.387	5.87	71.61	0.563
$SARIMAX(1,1,0)(0,0,1)_{12}$ [logRNGC]	4.402	Q = 13.813, sig. = 0.467	5.15	68.92	0.602
$SARIMAX(1,1,0)(1,0,0)_{12}$ [AT, logRNGC]	3.658	Q = 13.60, sig. = 0.628	3.95	54.59	0.628
SO_2_OLS_Regression [$AT_{(t-1)}$, $logRNGC_{(t-1)}$,$SO_{2(t-1)}$]	–	–	4.87	52.90	0.593

MAPE, mean absolute percentage error
[a] Identified models with best scores are bold.

parameter estimates ($p < 0.05$), arranged model equations of PM_{10} and SO_2 were obtained as follows:

$$\begin{aligned} PM_{10t} = 6.416 + PM_{10t-1} - 1.098(PM_{10t-12} + PM_{10t-13}) \\ - 0.098(PM_{10t-24} + PM_{10t-25}) - 0.541(logRNGC_{t-1} \\ - logRNGC_{t-2}) + 0.244(logRNGC_{t-12} - logRNGC_{t-13}) \\ - 0.309(AT_{t-1} - AT_{t-2}) - 0.646(AT_{t-12} - AT_{t-13}) \\ + (1 + 0.208B)\epsilon_t \end{aligned} \quad (9)$$

$$\begin{aligned} SO_{2t} = -1.172 + 0.871SO_{2t-1} + 0.129SO_{2t-2} \\ - 0.324(logRNGC_{t-1} - logRNGC_{t-2}) + 0.083(AT_{t-1} + AT_{t-2}) \\ + \epsilon_t \end{aligned} \quad (10)$$

Figures 5 and 6 show the short-term estimations of PM_{10} and SO_2 levels for 2013–2015, respectively. On the basis of RNGC and AT estimates, these visualizations indicated that the identified

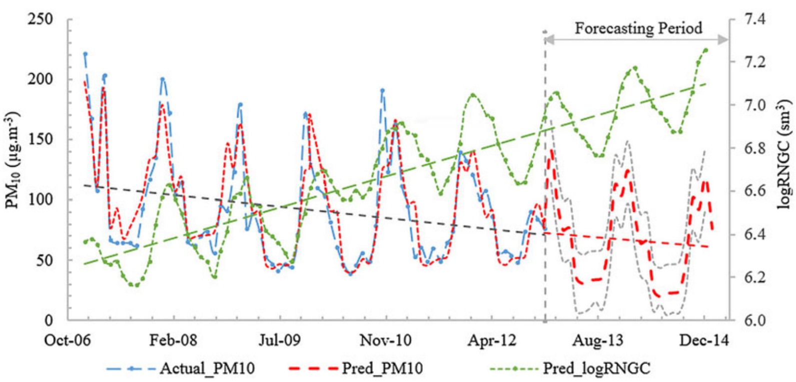

Figure 5. $SARIMAX(0,1,1)(1,1,0)_{12}$ model estimates for PM10 depending on the short-term predictions of RNGC and AT.

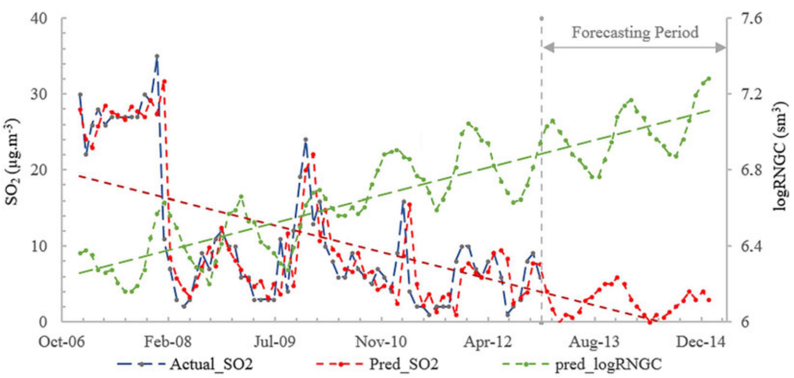

Figure 6. ARIMAX(1,1,0) model estimates for SO_2 depending on the short-term predictions of RNGC and AT.

Table 6. Summary statistics of short-term RNGC, PM_{10}, and SO_2 estimates for 2014–2015

Statistics	RNGC (sm^3) ARIMAX(1,1,2)			PM_{10} ($\mu g/m^3$) SARIMAX$(0,1,1)(1,0,0)_{12}$			SO_2 ($\mu g/m^3$) ARIMAX(1,1,0)		
	2013	2014	2015	2013	2014	2015	2013	2014	2015
Average	8.51E6	11.03E6	14.33E6	74	62	52	7	6	5
Winter	9.32E6	12.11E68	15.72E6	92	80	69	4	3	2
Summer	7.67E6	9.96E6	12.93E6	51	40	31	3	2	2
Maximum	12.10E6	15.71E6	20.41E6	141	124	118	11	9	8
Minimum	5.81E6	7.54E6	9.79E6	30	19	9	1	1	1
SD	2.17E6	2.86E6	3.66E6	38	35	31	3	3	2

SD, standard deviation

multivariate models represented the historical pattern acceptably well with moderately high values of R^2. However, in the previous time series analysis studies [10–14] mostly univariate models have been constructed and energy consumption and fuel shifting were not considered in forecasting air pollution at all. In this case, we concluded that multivariate time series models based on energy consumption and meteorological factors may be utilized in forecasting short-term air pollution with a reasonable accuracy according to times series forecasting plots.

Table 6 encapsulates the short-term projections by summarizing the statistics of identified TSMs. The estimated average monthly PM_{10} levels were in the range of 9–141 $\mu g/m^3$, while the yearly averages for 2013–2015 were estimated to be 74 ± 38, 62 ± 37, and 52 ± 31 $\mu g/m^3$, respectively. These figures meet national air quality guidelines, but not the annual standards of the U.S. Environmental Protection Agency, EU, and World Health Organization [3, 18]. On the contrary, short-term SO_2 estimates meet all standards given in several research papers [3, 18, 21] with significantly lower average monthly levels being estimated to be in the range of 1–11 $\mu g/m^3$.

4 Conclusion

The present study investigated short-term RNGC modelling and the abatement of air pollution from PM_{10} and SO_2 as a result of NG substitution, by using time series modelling techniques. The most significant parameter among others on the variations in RNGC was found to be AT that was fitted by a seasonal model. In the estimation of RNGC, the AT dependent ARIMAX(1,1,2) model yielded the best scores representing historical data and the trend in increase. Based on the short-term AT and RNGC estimates for 2013–2015, the SARIMAX$(0,1,1)(1,1,0)_{12}$ and ARIMAX(1,1,0) models produced the most satisfactory scores for short-term estimates of PM_{10} and SO_2 levels, respectively. The estimated PM_{10} levels indicated the abatement of fossil fuel air pollution to a level approximately three times lower than the average of the past six years. Although short-term PM_{10} levels meet the national air quality regulations, winter averages of PM_{10} remain high according to international standards. In contrast, the lower levels of SO_2 estimates meet all air quality standards. The results revealed that the estimated PM_{10} levels, rather than those of SO_2, can still pose health risks during the winter, despite increasing RNGC. The approach employed in this study resulted in a better understanding of short-term RNGC and its positive effect of air pollution by modelling, considering fuel shifting. However, further research is required to more accurately construct improved models for realistic estimates, and should include compilation of additional data such as daily fuel consumption statistics for the other fuel types in interest.

The author has declared no conflict of interest.

References

[1] L. K. Wang, N. C. Pereira, Y. Hung, Air Pollution Control Engineering, in *Handbook of Environmental Engineering*, Vol. *1*, Humana Press, Totowa, NJ **2004**, pp. 7–10.

[2] M. Kampa, E. Castanas, Human Health Effects of Air Pollution, *Environ. Pollut.* **2008**, *151*, (*2*), 362–367.

[3] A. C. Köne, T. A. Büke, Comparison for Turkish Provinces' Performance of Urban Air Pollution, *Renewable Sustainable Energy Rev.* **2012**, *16*, 1300–1310.

[4] A. D. Bhanarkar, P. S. Rao, D. G. Gajghate, P. Nema, Inventory of SO_2, PM and Toxic Metals Emissions from Industrial Sources in Greater Mumbai, India, *Atmos. Environ.* **2005**, *39*, 3851–3864.

[5] N. Ilten, A. T. Selici, Investigating the Impacts of Some Meteorological Parameters on Air Pollution in Balikesir, Turkey, *Environ. Monit. Assess.* **2008**, *140*, 267–277.

[6] C. Vahlsing, K. R. Smith, Global Review of National Ambient Air Quality Standards for PM_{10} and SO_2 (24 h), *Air Qual. Atmos. Health* **2012**, *5*, 393–399.

[7] G. M. Box, G. C. Jenkins, Reinsel, *Time Series Analysis: Forecasting and Control*, 3rd Ed, Prentice Hall, New Jersey **1994**.

[8] J. Kao, S. Huang, Forecasts Using Neural Network versus Box-Jenkins Methodology for Ambient Air Quality Monitoring Data, *J. Air Waste Manage. Assoc.* **2000**, *50*, (*2*), 219–226.

[9] J. Horowitz, S. Barakat, Statistical Analysis of the Maximum Concentration of an Air Pollutant: Effects of Autocorrelation and Nonstationarity, *Atmos. Environ.* **1979**, *13*, 811–818.

[10] T. Slini, K. Karatzas, N. Moussiopoulos, Statistical Analysis of Environmental Data as the Basis of Forecasting: An Air Quality Application, *Sci. Total Environ.* **2002**, *288*, 227–237.

[11] A. Kumar, P. Goyal, Forecasting of Daily Air Quality Index in Delhi, *Sci. Total Environ.* **2011**, *409*, 5517–5523.

[12] P. Goyal, A. T. Chan, N. Jaiswal, Statistical Models for the Prediction of Respirable Suspended Particulate Matter in Urban Cities, *Atmos. Environ.* **2006**, *40*, 2068–2077.

[13] M. H. Lee, N. H.A. Rahman, M. T. Surhatono, M. E. Latif, N. A.B. Nor, Kamisan, Seasonal ARIMA for Forecasting Air Pollution Index: A Case Study, *Am. J. Appl. Sci.* **2012**, *9*, (*4*), 570–578.

[14] L. Jian, Y. Zhao, Y. P. Zhu, M. B. Zhang, D. Bertolatti, An Application of ARIMA Model to Predict Submicron Particle Concentrations from Meteorological Factors at a Busy Roadside in Hangzhou, China, *Sci. Total Environ.* **2012**, *426*, 336–345.

[15] R. Modarres, A. K. Dehkordi, Daily Air Pollution Time Series Analysis of Isfahan City, *Int. J. Environ. Sci. Technol.* **2005**, *2*, (*3*), 259–267.

[16] H. K. Ozcan, Long Term Variations of the Atmospheric Air Pollutants in Istanbul City, *Int. J. Environ. Res. Public Health* **2012**, *9*, (*3*), 781–790.

[17] K. Kaygusuz, Energy Policy and Climate Change in Turkey, *Energy Convers. Manage.* **2003**, *44*, 1671.

[18] S. Erdogmuş, H. Aras, E. Koç, Evaluation of Alternative Fuels for Residential Heating in Turkey Using Analytic Network Process (anp) with Group Decision-Making, *Renewable Sustainable Energy Rev.* **2006**, *10*, 269–279.

[19] K. Kaygusuz, Energy Use and Air Pollution in Turkey, *Clean – Soil Air Water* **2007**, *35*, (*69*), 539–547.

[20] M. Koçak, C. Theodosi, P. Zarmpas, U. Ima, A. Bougiatioti, O. Yenigun, N. Mihalopoulos, Particulate Matter (PM10) in Istanbul: Origin, Source Areas and Potential Impact on Surrounding Regions, *Atmos. Environ.* **2011**, *45*, 6891–6900.

[21] T. Elbir, A. Muezzinoğlu, A. Bayram, Evaluation of Some Air Pollution Indicators in Turkey, *Environ. Int.* **2000**, *26*, 5–10.

[22] T. Elbir, A. Muezzinoglu, Estimation of Emission Strengths of Primary Air Pollutants in the City of Izmir, Turkey, *Atmos. Environ.* **2004**, *38*, (*13*), 1851–1857.

[23] G. Finz, G. Fronza, A. Spirito, Multivariate Stochastic Models of Sulfur Dioxide Pollution in an Urban Area, *J. Air Pollut. Control Assoc.* **1980**, *30*, (*11*), 1212–1215.

[24] A. Kadiyala, A. Kumar, Multivariate Time Series Models for Prediction of Air Quality Inside a Public Transportation Bus Using Available Software, *Environ. Prog. Sustainable Energy* **2014**, *33*, (*2*), 337–341.

[25] TURKSTAT, Online web service of Turkish Statistical Institute (TURKSTAT), Turkey **2012**.

[26] Düzce Municipality Strategic Plan for 2010-2014 (in Turkish), Düzce, Turkey **2010**.

[27] AKSA, AKSA Düzce-Eregli Gas Distribution Company (AKSA DERGAZ) Private Report, Düzce, Turkey **2012**.

[28] IEA-2012, World Energy Outlook 2012 Executive Summary, International Energy Agency (OECD/IEA), Paris **2012**.

[29] BOTAS, The Annual Sector Report-2011 (in Turkish), BOTAS Petroleum Pipeline Corporation, Ankara **2011**.

[30] H. H. Harman, *Modern Factor Analysis*, 2nd rev. Ed, University of Chicago Press, Chicago **1968**.

[31] G. D. Thurston, K. Ito, R. A. Lall, Source Apportionment of U.S. Fine Particulate Matter Air Pollution, *Atmos. Environ.* **2011**, *45*, (*24*), 3924–3936.

[32] M. Z. Ibrahim, R. Zailan, M. Ismail, M. S. Lola, Forecasting and Time Series Analysis of Air Pollutants in Several Area of Malaysia, *Am. J. Environ. Sci.* **2009**, *5*, (*5*), 625–632.

[33] L. Suganthi, A. A. Samuel, Energy Models for Demand Forecasting—A Review, *Renewable Sustainable Energy Rev.* **2012**, *16*, (*2*), 1223–1240.

[34] F. Taşpınar, N. Çelebi, N. Tutkun, Forecasting of Daily Natural Gas Consumption on Regional Basis in Turkey Using Various Computational Methods, *Energy Buildings* **2013**, *56*, 23–31.

[35] R. L.R. Salcedo, M. C.M. Alvim, C. A. Ferraz, F. G. Alves, Martins, Time-Series Analysis of Air Pollution Data, *Atmos. Environ.* **1999**, *33*, (*15*), 2361–2372.

[36] B. Billah, M. L. King, R. D. Snyder, A. B. Koehler, Exponential Smoothing Model Selection for Forecasting, *Int. J. Forecasting* **2006**, *22*, (*2*), 239–247.

[37] G. Ljung, G. Box, On a Measure of Lack of Fit in Time Series Models, *Biometrika* **1978**, *65*, 297–303.

[38] G. E. Schwarz, Estimating the Dimension of a Model, *Ann. Stat.* **1978**, *6*, (*2*), 461–464.

