Ground-level Ozone Modeling and Analysis in Selangor, Malaysia

Mohammed F.M. Abushammala[1], Mohammad K.Younes[2+], Abdul Amir H. Kadhum[3]

[1]Department of Civil Engineering, Middle East College, Knowledge Oasis Muscat, PB No 79, Al Rusayl124, Sultanate of Oman.
[2]Department of Civil Engineering, Philadelphia University, P.O.Box 19392, Amman, Jordan.
[3]Department of Chemical and Process Engineering, Faculty of Engineering and Built Environment, UniversitiKebangsaan Malaysia, Selangor, Malaysia

[+]Corresponding Author: myounes@philadelphia.edu.jo;PH: 00962 79 1368 173

Abstract:

Tropospheric ozone concentrations are negatively affect human health, the environment and agriculture activities.In addition, it is a key measure of air quality index which produced from the reaction of the emitted pollutants. This study aimed to develop Auto-Regressive Integrated Moving Average (ARIMA) model to predict daily ground-level ozone concentrations in Selangor, Malaysia. Average daily ozone concentration data were collectedfrom two mentoring stations in the state of Selangor; Kajang and Petaling Jaya, between 2005 and 2008 and used to construct the model. The results indicated that the ozone concentrations are daily altered with slow rise over the study period. In addition, it is affected by the previous concentrations where ARIMA (2,1,1) model is best representation of ground Ozone. Themean square errorfor the model is equal to 0.027 and model standardized residuals are within the accepted (± 2) confident interval which indicates the goodness of the suggested model. However, such model is applicable even though with lack of measurements of main pollutants.

Keywords: Ozone concentration, tropospheric ozone, ARIMA, Ozone modeling

Introduction:

In the recent years, air quality and pollution became an important public concern. Groundlevel ozone(O3) is negatively impacts human health and environment. It may lead to chronic respiratory infection and lunge inflammation. Furthermore, it is responsible for immunity system reduction, asthma, low biomass and plant production. Ozone is being considered as the most important secondary air pollutant, means it doesn't emitted directly to atmosphere. But it is produced as a result of photochemical reaction of various nitrogen oxides (NOx) with wide range of volatile organic compounds (VOC's) like solvent vapours and other hydrocarbons[1].However, the VOC's contribution to form ozone varies independence with their reactivity and chemical constitutions [2].These primary pollutants, that form ozone, are usually emitted as a result of anthropogenic activities like transportation and industry. On the other hand, ozone is destroyed by oxidation action of additional reactant (D). Ozone formation and destroying can be expressed as the following reaction.

In addition, many other factors affect ground ozone formation such as metrological condition, sunlight (UV) and primary pollutants concentrations, for instance ozone dispersion as other gaseous pollutant mainly affected by wind speed and velocity.

$$NOx + VOCs + UV \longrightarrow O3 \quad \ldots.. (1)$$

$$O3 + D \longrightarrow DO + O2 \quad \ldots..(2)$$

Generally, prediction can be classified into (1) qualitative techniques based on expert opinion and/or personal judgment and (2) quantitative techniques based on mathematical models. Quantitative forecasting is further classified into time series or causal forecasting. Causal forecasting searches for relationship between an input variables and the desired output. While time series forecasting uses past data to predict the future one. Time series forecasting is flexible technique, and its implementation does not require much data. The main limitation of time series analysis is the lack of deterministic cause [3]. However, building the model in dependence on large number of stochastic events may overcome this limitation.

Ozone forecasting is complex and linked with various factors that control the ozone formation. In addition, the availability and accuracy of such factor is questionable specially in developing countries.Linear regression has a long ozone forecasting history [4], it's main advantages is the relatively high accuracy with moderate to low cost . However, the developed linear regression models are simple and limited in determining some variables. In addition, ozone formation is complex and non linear. For instance, photochemical ozone formation rises in presence of NOx but it is less sensitive to VOC existence [5].

Comrie, 1977, compared the performance of Artificial Neural Network (ANN) and multivariate regression to model and forecast ozone concentration under various climatic and ozone regimes. His work proved applicability of NN implementation to forecast the ozone[6].Furthermore, implementation of the artificial neural network to predict the ozone concentration at the next day in Athens at summer time has been tested by [7]. The results show that the ANN performance is better than liner regression. However, the unknown working principle or "black-box" of neural network limits its wide application and acceptance to predict the Ozone level.Fuzzy logic and ANN were used to predict the possibility of high ozone level that may beoccurred at the next day [8]. Ozone predicting using photochemical models result in acceptable level of accuracy. But implementing of such model requires large and complicated inputs data and sophisticated processing regime. Moreover, the variations of ozone concentration independence of geographical, atmospheric and socioeconomical aspects were investigated by [9].

However, ozone concentration is partially dependent on previous concentration thus it formsa kind of serial correlation, thus time series forecasting is implemented for ozone forecasting. This work aims to build a time series ARIMA model to forecast ozone concentration, which will help ozone control strategise and reduce the injuries and damaged due to ground ozone. The Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) will be used to determine the structure of ARIMA model[10] and the RMSE will evaluate its performance.

Methodology:

Study Area

Selangor state is the most populated state in Malaysia, according to Malaysian Department of Statistics (2016) the population in Selangor state is 6.3 million, Selangor area is 7,930 km^2.

Selangor is the richest state in Malaysia and has the largest industrial and commercial progress thus it witness the largest volume of air pollution. In addition, the rapid increases of registered vehicles in Malaysia in the last few yearsworsen the air quality. For instance according to the department of road transportation, the registered cars were only 10.5 million in 2000 with (0.45 car per capita) and became 25.1million registered vehicles in whole Malaysia (0.86 car per capita) by 2015.Two Ozone monitoring stations were chosen to represent the air quality in Selangor State. The first one is at Kajang city , while the second one is at Klang city. Kajang is a hub city located around 21 km east of Kuala Lumpur, due to its unique location the population increases rapidly, the growth rate is around 9% per year. On the other hand, Klang is the former capital of Selangor and it has Port Klang, themain port of Malaysia.

Data Collection and Modeling:

Time series forecasting is multidisciplinary scientific tools that is used to solve the prediction problems. It needs only the historical observations of the required variables, so its implementation is easy and flexible [11-13] . ARIMA method first presented by Box and Jenkin in 1976 [11]. The general equation of the successive difference at d^{th} difference of X_t is defined by

$$\Delta^d X_t = (1 - B)^d X_t \quad \text{.....(3)}$$

Where
d is the difference order and is usually 1,2.
B is the backshift operator.

The successive difference at one time lag equals to

$$\Delta^1 X_t = (1 - B) X_t = X_t - X_{t-1} \quad \text{.....(4)}$$

For purpose of this work, the general ARIMA (p, d, q) is briefly expressed as the follows:

$$\Phi_p(B) W_t = \theta_q(B) e_t \quad \text{.....(5)}$$

Where

$\Phi p(B)$ is an autoregressive operator of order p

θq (B) is a moving average operator of order q

and $W_t = \Delta dX_t$[14]

Hourly data of tropospheric ozoneconcentration, from Kajang and Klang stations, for a period of five years between 1 January 2005 and 31 March 2010 were obtained from the Department of Environment Malaysia. Then the data was validated and a daily average values were calculated to find an average ozone concentration[15].Table (1) describes the collected data. ARIMA modeling was developed using Matlab R2012a (7.14) software that also was used to prepare the data and find the ACF, the PACF and the first deference.

Table 1 Summary of ground Ozone data descriptive statistics

Variables	**Value**
No of readings	1,916
Min.	3 $\mu g/m^3$
Max.	110 $\mu g/m^3$
Mean	17 $\mu g/m^3$
Median	16 $\mu g/m^3$
Standard Deviation	7.7
Range	107

Results and Discussions:

The collected data is tested against stationery test. However, since the data is none- stationary the first difference was applied to stationeries the data. The stationerized data is shown in Figure (1), it is obvious that the data after stationary fluctuate around the zero thus the first difference is enough and no need to go for further differences. The results of ACF and the PACF tests are shown in Figure (2). The ACF is responsible for determining the Moving Average (MA) part of ARIMA model; it shuts off after first lag. While the PACF is responsible for Autoregressive (AR) and it shuts off after five lags. However, the model shows reasonable results (statistically significant results) after the second lag of PACF, thus the final model structure assumed to be ARIMA (2, 1, 1). However, the model can be expressed by the following equation

$$y_t = 0.0046 + 0.3468y_{t-1} + 0.0190y_{t-2} + \varepsilon_t - 0.9315\varepsilon_{t-1} \ldots..(6)$$

Where, y_t is the daily average of ground ozone $\mu g/m^3$ at day (t).

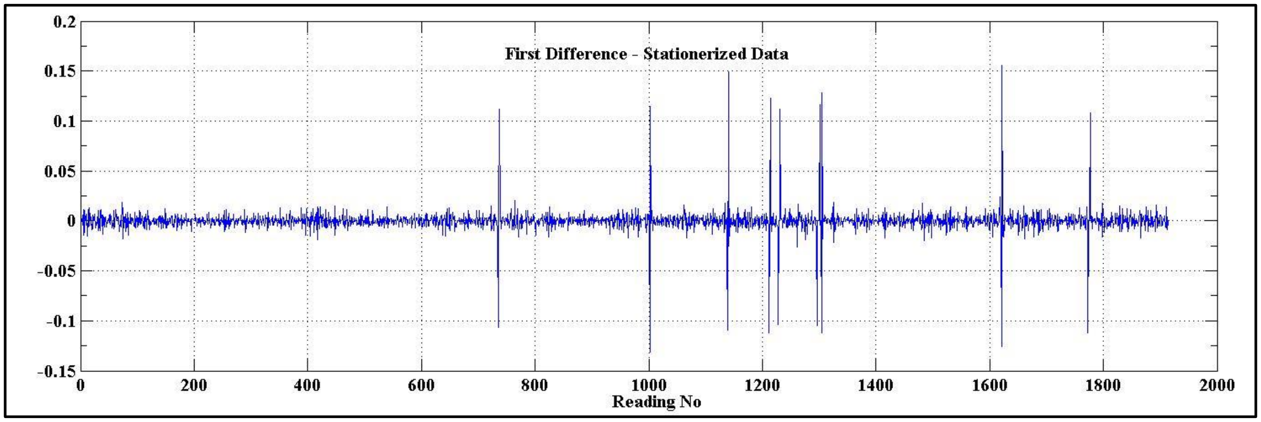

Figure 1 First difference of Original ground Ozone data

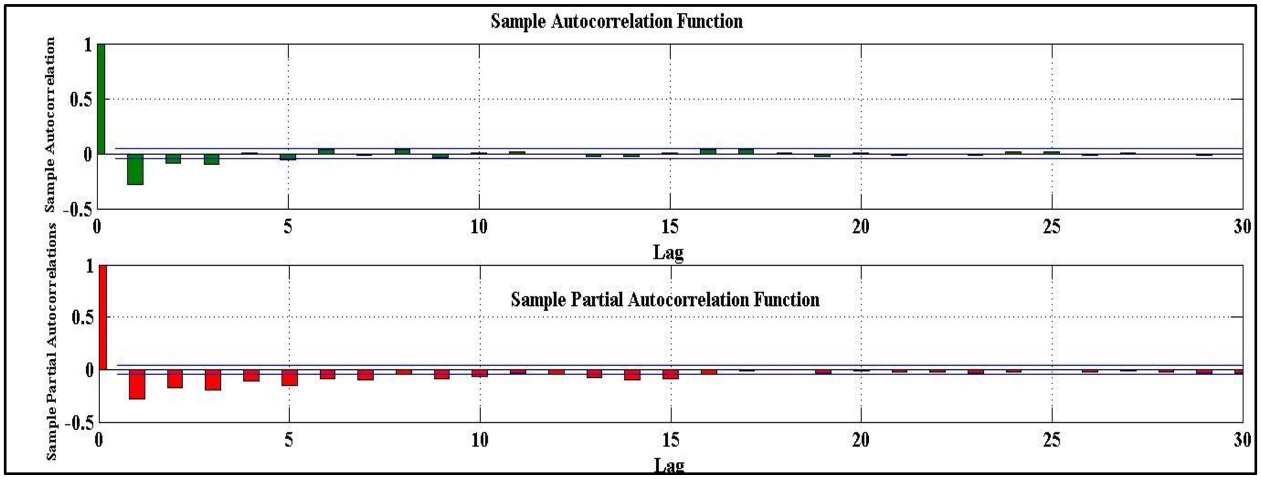

Figure 2 ACF and PACF for first differenced data

Table (2) shows the model performance

Parameter	Value	Standard Error	T statistics
Constant	0.0046	0.0012	0.3318
AR{1}	0.3468	0.0107	32.3315

AR{2}	0.0190	0.0129	1.4787
MA {1}	-0.9315	0.0096	-97.5144
Variance	0.0039	0.0029	134.8820

Figure (3) shows the model performance for 100 days forecast with expected maximum and minimum scenario with 95% interval. The upper and lower limits have been developed based on the mean square error. However, the expected values will be in the range between 24 – 27 µg/m^3 of ground Ozone which is far less than the guideline value for ozone level for an 8-hour daily average that is 100 µg/m^3. In addition, even the maximum ozone value in the upper limit is 46µg/m^3 that is almost half of the guideline value.On the other hand, the general trend of ground ozone value is increasing with time, which highlights the demand for required actions to reduce its impact on human and environment health.

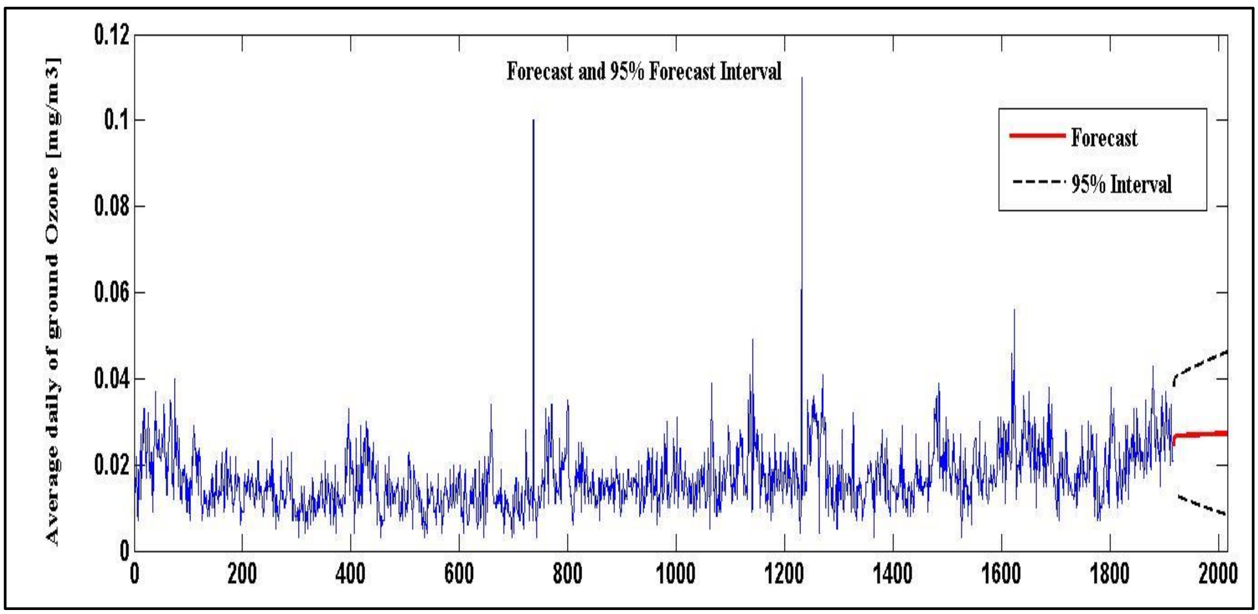

Figure 3 Daily ground Ozone forecast for 100 days

To examine the model goodness standardized residuals were calculated and shown in figure (4). Most of the residual values lay between ± 2, thus the standardized residual may be characterized as normally distributed residuals. However, the positive and negative values of the standardized residuals are another indication for model goodness, it means that the predicted values sometime are more than the original ones and another time they are less than it. Although there are some residuals falls beyond ± 2 but these arelimited number and still within the accepted 95% confident interval. Figures 5 and 6 show ACF and PACFfor sixteen lags,respectively. However, since all the expressed lags are within the critical value, thus it can be concluded that the residuals are uncorrelated. In addition, the mean square error for the suggested model is 0.027 so the model performance is good and accepted.

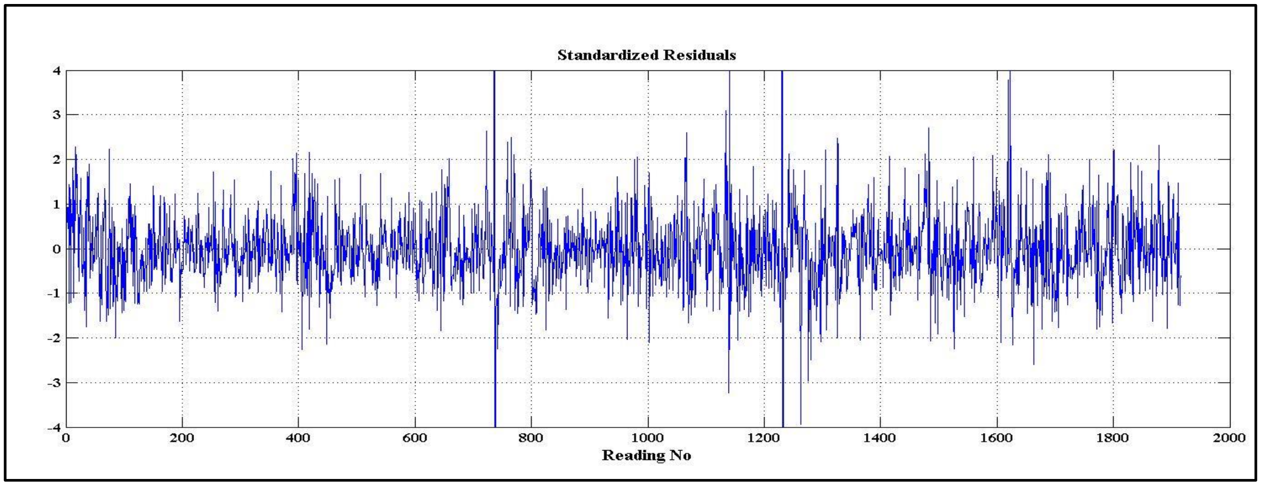

Figure 4 Standardized residuals

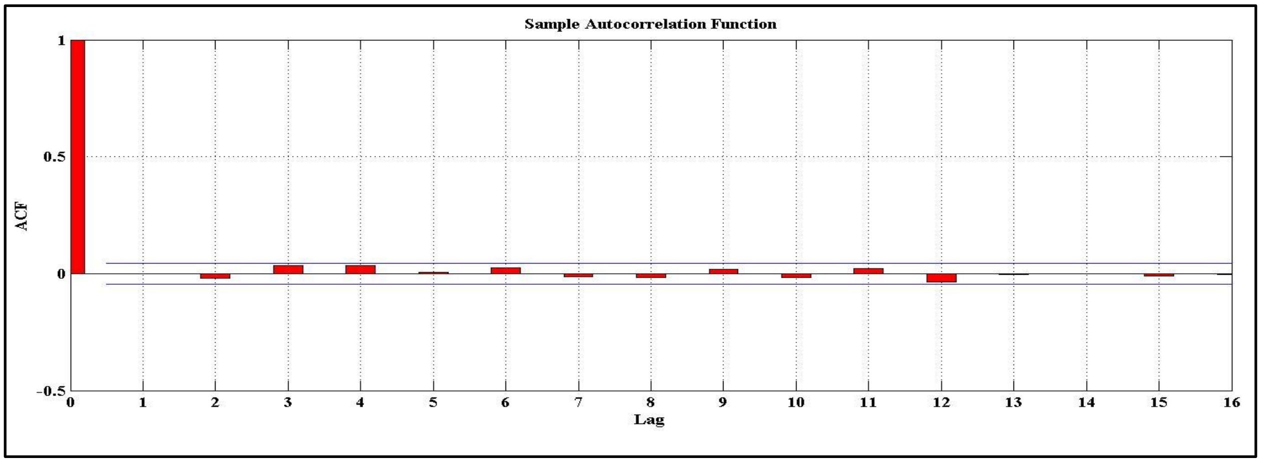

Figure 5 Residual ACF

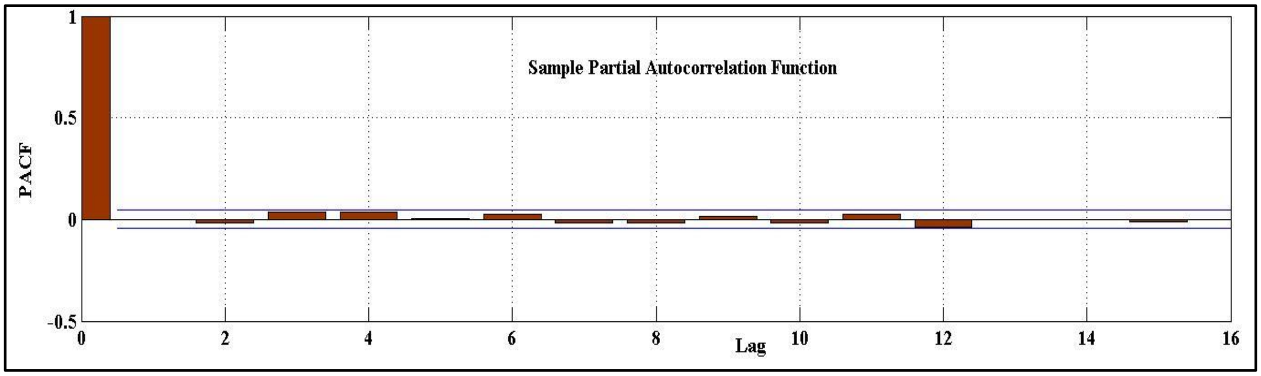

Figure 6 Residual PACF

Finally, comparison of ARIMA model with Monte Carlo simulation for the next one hundred readingsis shown in Figure (7). Although, Monte Carlo simulation represent a relatively larger fluctuations, but the ARIMA model performance and Monte Carlo simulation mean are virtually indistinguishable.Furthermore, there are slight discrepancies between the theoretical 95% forecast intervals and the simulation-based 95% forecast intervals.

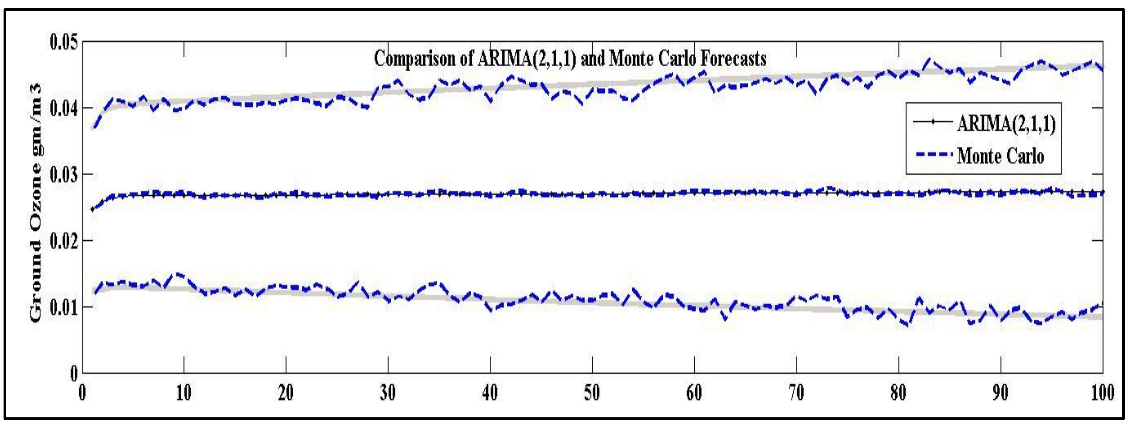

Figure 7 Comparison of ARIMA Forecast Vs Monte Carlo simulation model.

Conclusions:

Generally speaking, ground ozone is an important air pollution index. Due to recent increases in industrial and vehicles transportation in Malaysia, the level of ground ozone is characterized by fluctuatingand increasing trends. Time series forecasting is convenient especially when not enoughdata is available. The average predicted daily ground ozone is between 24-27μg/m^3 which is less than the WHO guideline (100μg/m^3) but the issue is increasing trend of ground ozone. However, considering this fact isessential for sustainable and proper planning especially in the field of transportation. Finally, testing modelgoodness using mean square error and the standardized residuals shows that the model residual are non-correlatedand normally distributed.

Acknowledgments:
The authors acknowledge the support provided by The National University of Malaysia, Philadelphia University and Middle East College for their support to complete this research.

References:

[1] U. Im, R. Bianconi, E. Solazzo, I. Kioutsioukis, A. Badia, A. Balzarini, *et al.*, "Evaluation of operational on-line-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2. Part I: Ozone," *Atmospheric Environment,* vol. 115, pp. 404-420, 2015.

[2] A. Melkonyan and P. Wagner, "Ozone and its projection in regard to climate change," *Atmospheric Environment,* vol. 67, pp. 287-295, 3// 2013.

[3] S. Makridakis, S. C. Wheelwright, and R. J. Hyndman, *Forecasting methods and applications*: John Wiley & Sons, 2008.

[4] N. V. Balashov, A. M. Thompson, and G. S. Young, "Probabilistic Forecasting of Surface Ozone with a Novel Statistical Approach," *Journal of Applied Meteorology and Climatology,* vol. 56, pp. 297-316, 2017.

[5] I. Mavroidis and M. Ilia, "Trends of NO x, NO 2 and O 3 concentrations at three different types of air quality monitoring stations in Athens, Greece," *Atmospheric environment,* vol. 63, pp. 135-147, 2012.

[6] A. C. Comrie, "Comparing neural networks and regression models for ozone forecasting," *Journal of the Air & Waste Management Association,* vol. 47, pp. 653-663, 1997.

[7] A. Chaloulakou, M. Saisana, and N. Spyrellis, "Comparative assessment of neural networks and regression models for forecasting summertime ozone in Athens," *Science of the Total Environment,* vol. 313, pp. 1-13, 2003.

[8] J.-S. Heo and D.-S. Kim, "A new method of ozone forecasting using fuzzy expert and neural network systems," *Science of the Total Environment,* vol. 325, pp. 221-237, 2004.

[9] A. Santurtún, J. C. González-Hidalgo, A. Sanchez-Lorenzo, and M. T. Zarrabeitia, "Surface ozone concentration trends and its relationship with weather types in Spain (2001–2010)," *Atmospheric Environment,* vol. 101, pp. 10-22, 1// 2015.

[10] J. Navarro-Esbrı, E. Diamadopoulos, and D. Ginestar, "Time series analysis and forecasting techniques for municipal solid waste management," *Resources, Conservation and Recycling,* vol. 35, pp. 201-214, 2002.

[11] M. Khashei and M. Bijari, "A novel hybridization of artificial neural networks and ARIMA models for time series forecasting," *Applied Soft Computing,* vol. 11, pp. 2664-2675, 2011.

[12] S. S. Chung, "Projecting municipal solid waste: the case of Hong Kong SAR," *Resources, Conservation and Recycling,* vol. 54, pp. 759-768, 2010.

[13] L. Xu, P. Gao, S. Cui, and C. Liu, "A hybrid procedure for MSW generation forecasting at multiple time scales in Xiamen City, China," *Waste management,* 2013.

[14] M. K. Younes, Z. Nopiah, N. E. A. Basri, H. Basri, M. Shitan, L. S. Lee*, et al.*, "Medium term municipal solid waste generation prediction by autoregressive integrated moving average," in *AIP Conference Proceedings*, 2014, pp. 427-435.

[15] W. H. O. WHO, "WHO Air quality guidelines for particulate matter,ozone, nitrogen dioxide and sulfur dioxide Global update 2005, Summary of risk assessment," pp. 15-17, 2006.