EnvironmentAsia
The international journal published by the Thai Society of Higher Education Institutes on Environment

Available online at www.tshe.org/EA
EnvironmentAsia 8(1) (2015) 53-60

An Analysis into the Temporal Variations of Ground Level Ozone in the Arid Climate of Makkah applying k-means Algorithms

Said Munir [a], Turki M. Habeebullah [a], Atef M.F. Mohammed [a], Essam A. Morsy [a], Abdel Hameed A.H. Awad [a], Abdulaziz R. Seroji [a] and Ibrahim A. Hassan [b]

[a] *The Custodian of the Two Holy Mosques Institute for Hajj and Umrah Research, Umm Al-Qura University, Makkah, Saudi Arabia*

[b] *Faculty of Science, Environmental Studies, Centre of Excellence in Environmental Studies, Air Pollution, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

Analysis of temporal variations helps identify the time when an air pollutant is more likely to exceed the air quality standards. This paper characterises the temporal variations of ground level ozone (O_3) in Makkah. In addition to graphical presentations, like time variation plots, this paper applies k-means algorithm to test its applicability for analysing the diurnal cycles of O_3. Diurnal, weekly and seasonal variations in O_3 concentrations are analysed using data for 2012. O_3 data for 1997 - 2007 are used to investigate the annual cycle of O_3. The average annual cycle shows the highest concentration in September and lowest in December. O_3 concentrations are lower in colder months probably due to lower solar radiation levels, whereas unexpectedly highest O_3 concentrations are observed in September. O_3 levels decline in the hottest months - June and July probably due to chemical and biophysical feedbacks. Furthermore, O_3 diurnal cycles are clustered into 4 and 12 subgroups using k-means algorithm. The clusters are considerably different from the monthly and seasonal diurnal cycles, probably because of the anomalies found in various months or seasons that biases the average cycles. These anomalies are addressed by k-means algorithm putting them into a separate group.

Keywords: ground level ozone; temporal variations; k-means algorithm; Makkah

1. Introduction

A decrease or increase in air pollution concentration is the result of an imbalance between air pollutants production rates and air pollutants removal rates (Andersson *et al*., 2006). Meteorological parameters (e.g., temperature, wind, relative humidity) play a vital role in photochemical ozone (O_3) formation, and influence the transport, dispersion and chemical reactions of air pollutants (Andersson *et al*., 2006; Baur *et al*., 2004; Duenas *et al*., 2002). Substantial variations in meteorological conditions can exert a large impact on O_3 concentrations and often mask long term trends in O_3 concentrations (Duenas *et al*., 2002; Gardner and Dorling, 1999). O_3 concentrations exhibit typical diurnal, weekly and annual cycles due to changes in meteorological parameters, O_3 precursors (e.g., nitrogen oxides (NOx), hydrocarbons (HCs), and carbon monoxide (CO)) and sinks (e.g., dry deposition or scavenging by nitric oxide (NO) (Hassan *et al*., 2013; AQEG, 2009).

Previously, temporal variations of ground level O_3 have been characterised by several researchers (e.g., Coyle *et al*., 2002; Jenkin *et al*., 2002; AQEG, 2009; Hassan *et al*., 2013). O_3 concentrations in rural areas show a mid-afternoon maximum and night-time minimum (Coyle *et al*., 2002). The afternoon peak in O_3 concentration is produced by photochemical O_3 production and the descending of O_3 from the free troposphere. The former is caused by NOx and HCs reactions in the presence of solar radiation, whereas the latter is caused by turbulent mixing induced by both wind shear and thermal convection (AQEG, 2009). In urban areas, the diurnal cycle of O_3 shows different characteristics and is controlled by NO, which reacts with O_3 and produces nitrogen dioxide (NO_2). Therefore, in urban areas the minimal O_3 concentrations are normally observed during the morning and evening traffic peak hours, while the maximum concentrations are observed at night (due to low traffic and hence less titration by NO) and in the mid-afternoon (due to photochemical O_3 formation) (Coyle *et al*., 2002; Munir *et al*., 2012).

The meteorological conditions associated with anticyclones, such as high solar radiation, high temperature, low wind speed and low rainfall are favourable for tropospheric O_3 formation, and these conditions are prevalent in Saudi Arabia (Hassan *et al*., 2013). Therefore, O_3 concentrations exceed the air quality standard of 120 μg/m^3 8 hour running mean set by WHO and European Union (Munir *et al*., 2013). Munir *et al*. (2013) analysed long term historic trends of several air pollutants, including O_3 and demonstrated a significant positive trend over the study period (1997 - 2012). Kurokawa *et al*. (2009) reported an increasing O_3 trends in Japan, and stated that the increasing trend of boundary layer O_3 was caused by the recent increase of anthropogenic precursor emissions in East Asia, especially in China.

It is vital to analyse O_3 concentration, assess its negative impacts, and know its temporal variations. The analysis of temporal variations helps identify the time when air pollutants are more likely to exceed the air quality standards. O_3 temporal variations are not well characterised in Makkah, Saudi Arabia. Therefore, the purpose of the present study is to determine the diurnal, weekly and seasonal cycles of O_3 in Makkah city, and analyse the various factors responsible for its temporal variations. Moreover, this paper investigates the potentiality of k-means algorithms for temporal analysis of O_3 concentrations which provide further insight into the temporal variations O_3.

2. Methodology

Air quality data of O_3 for year 2012 were analysed. The data were collected from two continuous monitoring stations, the Presidency of Meteorology and Environment (PME) and Masfalah monitoring stations. The monitoring stations are situated near the Holy Mosque (Al-Haram) in Makkah (Fig. 1). O_3 data during 1997 - 2007 are also used in this paper to characterise the annual variations. These are both continuous monitoring stations and measure the concentrations of several air pollutants and meteorological variables. The monitoring stations were previously described by Munir *et al*. (2013) and Habeebullah (2013). PME is a background monitoring site located inside the Holy Mosque, whereas Masfalah is a roadside monitoring site.

In this paper the temporal variations of ground level O_3 are analysed using time variation plots (Carslaw and Ropkins, 2012) and k-means clustering algorithm (MacQueen, 1967). K-means clustering is a method of cluster analysis that divides 'n' observations into 'k' clusters (sub-group). Each observation in the cluster

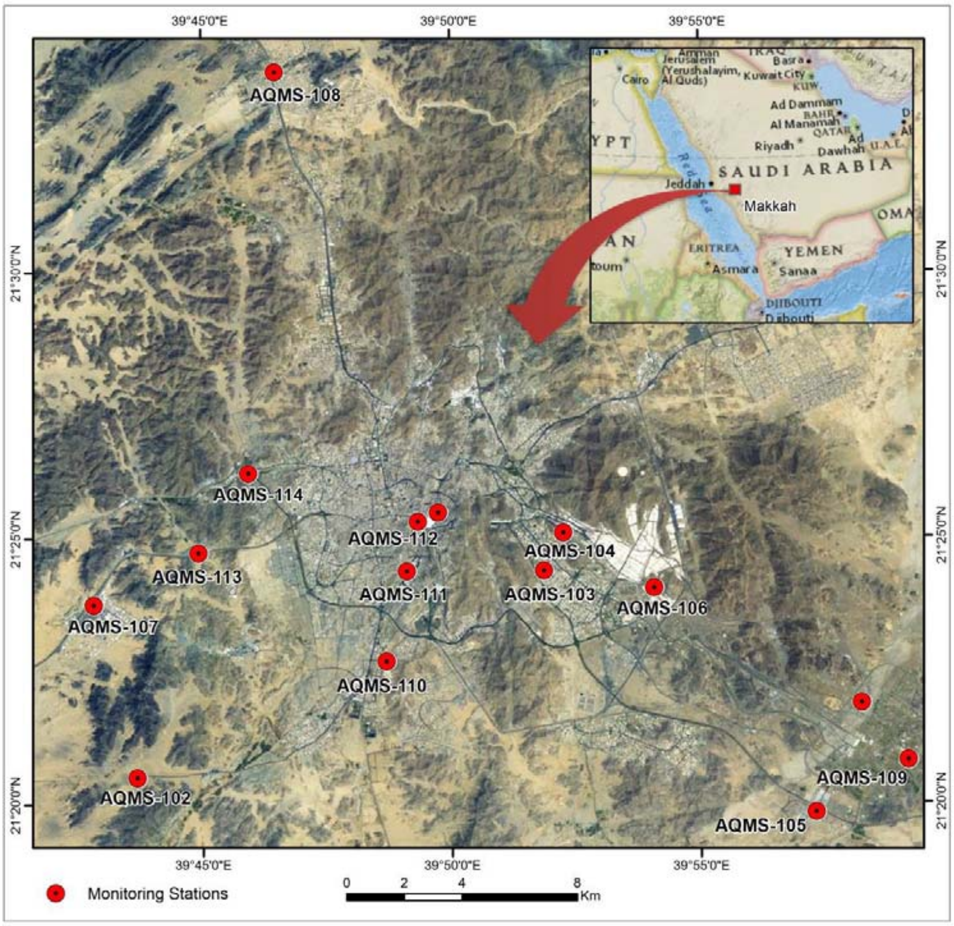

Figure 1. Map of the air quality and meteorological monitoring sites in Makkah, Saudi Arabia, where AQMS 112 represents the PME site and AQMS111 represents the Masfalah site.

belongs to the cluster with the nearest mean. The algorithm used is called k-means algorithm which is an iterative refinement technique. K initial means are randomly selected from a given dataset. The better option is to select them as far away from each other as possible. In the next step, each observation is associated with the nearest mean to create k clusters. For each k cluster new k centroids (barycentres) are calculated. The centroid of each of the k-clusters becomes the new means. Creating k clusters and calculating new mean for them is repeated until no more changes occur (convergence has reached). The algorithm aims at minimising a squared error function (the sum of squared distances to the cluster centres). A simple approach to find the optimal number of clusters is to have multiple runs with different k classes and choose the best one. It is important to note that increasing k, results in smaller error function but also increases risk of overfitting. In this study hourly O_3 data were first converted into a 24 columns format and then k-means algorithm was applied.

Statistical data analysis was carried out in the statistical software R programming language (R Development Core Team, 2012), and its package openair (Carslaw and Ropkins, 2012).

3. Results and Discussion

Fig. 2 shows time variation plots of O_3 at the PME and Masfalah monitoring sites. The annual average of O_3 concentration is higher at the PME site (70.46 ug/m^3) than the Masfalah site (59.14 ug/m^3). O_3 concentrations exhibit a typical diurnal cycle, lowest during the morning, especially during the peak hours and highest during the afternoon. Higher O_3 concentration during the day is expected due to the fact that O_3 is a secondary air pollutant, which is produced by the photochemical reaction of HCs and NOx in the atmosphere. During the night-time due to the lack of solar radiation photochemical O_3 formation no longer takes place, and O_3 concentration is further decreased by dry deposition (the accumulation of O_3 as it comes into contact with soil, water or vegetation on the earth's surfaces) and scavenging (consuming) by NOx species ($O_3 + NO \rightarrow NO_2 + O_2$). NO concentrations are high during the morning traffic peak hours and probably that is why O_3 concentrations reach the lowest level, as NO is negatively correlated with O_3 (Munir *et al.*, 2012; Jenkin, 2004). After the sun rise due to photochemical O_3 formation, the concentrations of O_3 increase and reach the highest levels in the afternoon. This typical O_3 cycle is in agreement with previous studies (e.g., Munir *et al.*, 2012; Coyle *et al.*, 2002; AQEG, 2009).

On weekly basis (Fig. 2, bottom-right), the highest O_3 concentration was observed on Friday at both monitoring sites. In Makkah in 2012 the weekend was on Thursday and Friday (remember: weekend was changed to Friday - Saturday in 2013). Friday observes less road traffic than the other days. This results in low emission of NO, which is an O_3 scavenger and hence O_3

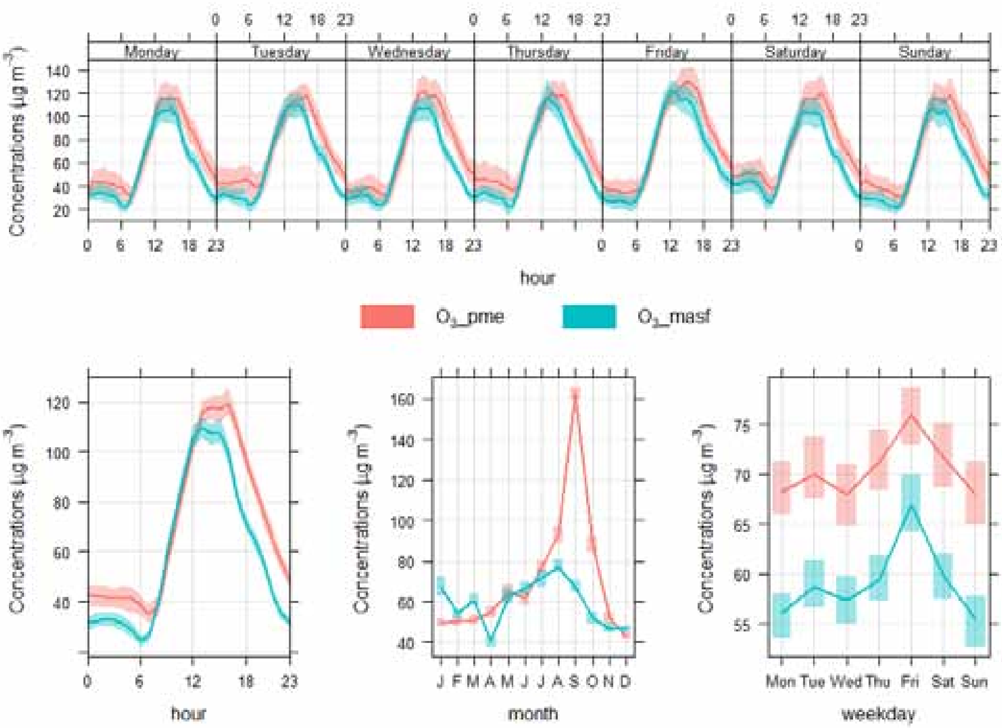

Figure 2. Time variation plots of O_3 concentrations (μg/m^3) at the PME and Masfalah monitoring sites in the year 2012.

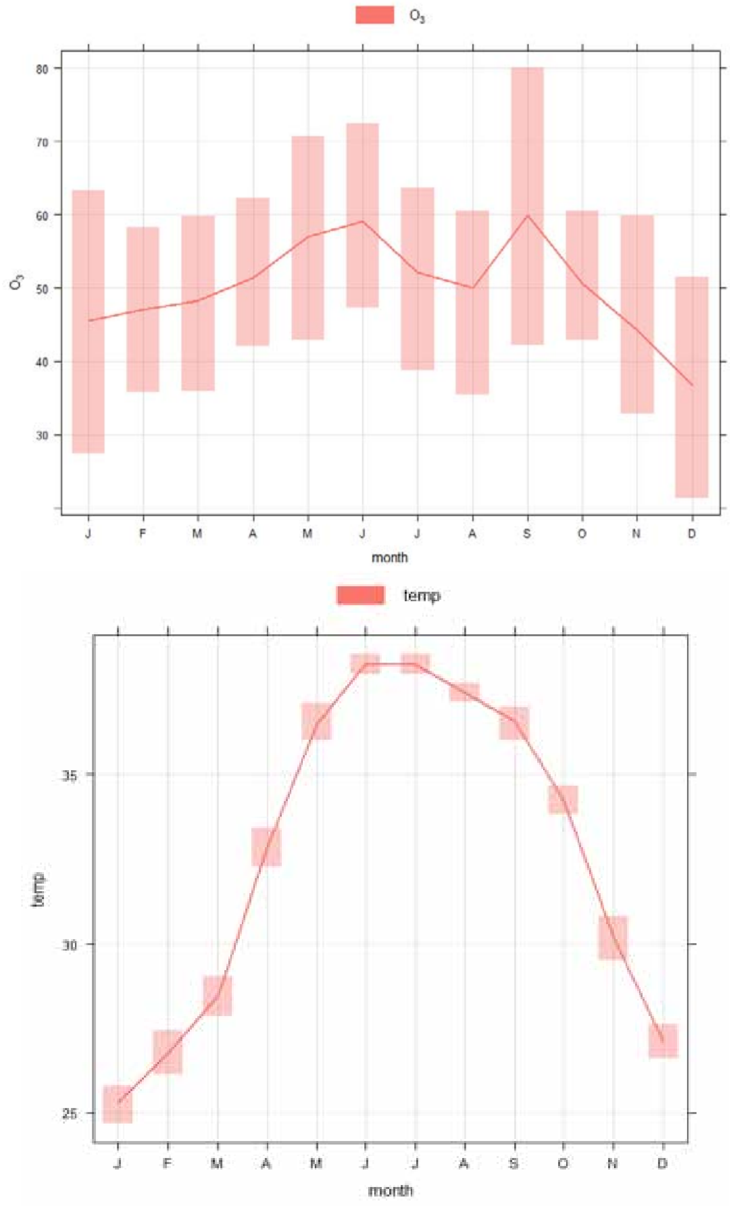

Figure 3. Average annual cycle of O_3 concentrations ($\mu g/m^3$) (top) and temperature (^oC) (bottom) at the PME monitoring site from 1997 to 2007.

levels are generally higher on weekends than working days. Higher O_3 level during weekend is a well known phenomenon and is referred to as 'ozone weekend effect (OWE)' (Jenkin, 2008). Compared to other weekdays, O_3 levels are higher on Saturday, this may be attributed to the carry over effect of O_3 from the weekend.

The annual cycle of O_3 is shown in Fig. 2 (middle-bottom), where the highest O_3 concentrations are shown in August (about 80 $\mu g/m^3$) and September (about 160 $\mu g/m^3$) at Masfalah and PME sites, respectively. At the PME site O_3 levels are lower in colder months and higher during the hotter months. Furthermore, the average annual cycle of O_3 for 11 years (1997 - 2007) (Fig. 3, top) shows the highest concentration in September and lowest in December. Overall, O_3 levels are lower in colder months (Figs. 2 and 3), which is expected due to lower solar radiation levels in these months. Interestingly, the highest O_3 concentrations are observed in September when the temperature levels are relatively low and not in June and July when the temperature levels are high (Fig. 3, bottom). This might show that in June and July the temperature level is too high in Makkah (hourly average temperature reaches as high as 50^oC as reported by Munir, 2014) and O_3 level starts decreasing. Steiner *et al*. (2010) have reported that low levels of O_3 at the extreme temperature are due to chemical and biophysical feedbacks. More recently, Munir (2014) has analysed the negative O_3 - temperature slope at extremely high temperature (> 42^oC) in Makkah and concluded that O_3 levels decrease at extremely high temperature probably due to reduction in the levels of O_3 precursors, such as NO_2 and total hydrocarbons.

3.1. Diurnal cycle of O_3 using k-means clustering

In this paper k-means clustering algorithm is used to divide the diurnal cycle of O_3 into 4 and 12 clusters. The reason for choosing 4 and 12 clusters is to make them comparable with the diurnal cycles of O_3 during various months and seasons of the year, where spring (March - May); summer (June - August); autumn (September - November); and winter (December - February) (Carslaw and Ropkins, 2012).

Fig. 4 (top) shows the diurnal cycle of O_3 during various seasons, where autumn and winter exhibited the highest and lowest O_3 levels, respectively. Furthermore, in the winter and summer, O_3 shows the highest level at about 14:00 hour, whereas in spring and autumn the highest level is observed at about 17:00 hour (Fig. 4). Fig. 4 (bottom) shows the results of k-means clustering, dividing the diurnal cycles into 4 clusters. The clusters are significantly different from the seasonal diurnal cycles, especially cluster 2 which shows considerably higher levels of O_3, even during the morning and night times when generally low levels of O_3 are expected. Cluster 1 and 3 are somewhat like winter and spring cycles, however the peak times are different. Furthermore, cluster 4 more or less appears to be like the diurnal cycle in autumn, however the morning levels are much higher in cluster 4. This might show that in each season the diurnal cycle of O_3 exhibits considerable variation which is probably caused by the day to day variations in meteorological conditions and the amount of O_3 precursors. This justifies the application of clustering approach for investigating O_3 diurnal cycle which provides further insight into the diurnal cycles of O_3. It is common to observe different cycles within the same season or the same cycles during different seasons. These anomalies can affect the

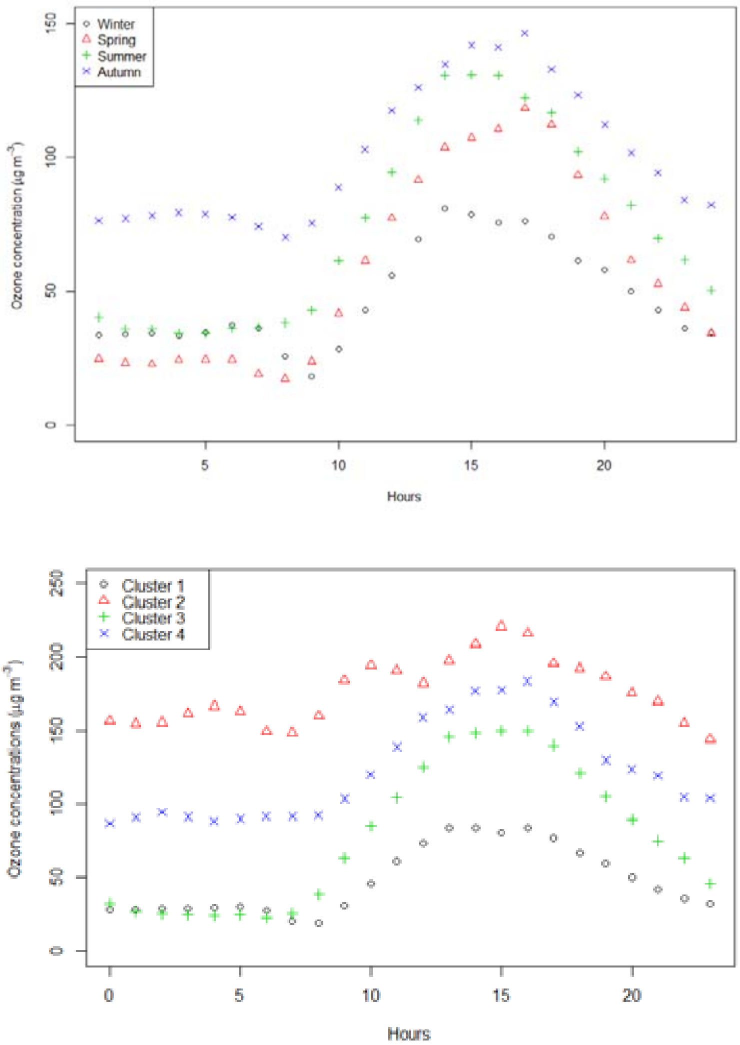

Figure 4. Diurnal cycles of O_3: during various seasons (winter, spring, summer, autumn) (top); and using k-means clustering approach (bottom) at the PME station for year 2012.

average cycle of O_3 in a particular season, therefore in addition to just averaging the diurnal cycles, there is a need for a methodologies that can separate these types of anomalies. This is successfully done by the k-means clustering algorithm.

Fig. 5 (left) shows the diurnal cycles of O_3 during different months of the year (2012). September shows the highest, whereas December shows the lowest O_3 concentration. This is worth mentioning that the difference in O_3 concentrations during day and night varies from month to month, e.g., in the early morning October shows higher O_3 levels than August, whereas the opposite is true in the afternoon. Similar trend is shown by May and December, where O_3 level is much lower during December in the afternoon. O_3 photochemical formation and its concentration in the atmosphere are directly related to the levels of O_3 precursors, solar radiation and temperature as explained in Munir *et al*. (2012). Fig. 5 (right) shows 12 clusters produced by k-means clustering. The 12 clusters are different from the monthly diurnal cycles shown in Fig. 5 (left). The cluster 3 is somewhat like September, however cluster 4 is totally different from August (or any other month). Likewise, cluster 11 showing the lowest O_3 concentration, is different from the month of December, which also observes lowest O_3 levels. Cluster 9 shows the greatest difference between morning and afternoon, whereas cluster 7 shows the lowest variation between morning and afternoon. Cluster 9 would probably represent August or May when photochemical O_3 formation is highest during the afternoon due to the high amount of solar radiation, whereas cluster 7 probably represents a cold cloudy day when little photochemical O_3 formation takes place. However these diurnal cycles are not shown in the 12 diurnal cycles representing each month, where such distinct behaviour disappears due to the averaging of 30 days or so for each month. This analysis shows that apart from seasonal and monthly variation in the diurnal cycles of O_3 there are some mixed days that show different characteristics within the same season or month. Therefore probably cluster analysis is better suited for analysing ozone diurnal cycle than just averaging over a certain period of time.

Nitrogen oxides (NOx), which is the sum of nitric oxide (NO) and nitrogen dioxide (NO_2), play an important role in controlling the levels of O_3. NOx is considered one of the precursors of O_3 that lead to the formation of O_3. On the other hand freshly emitted NOx, especially NO is a sink of O_3. NO reacts with O_3 and converts it to NO_2 and atomic oxygen and, therefore is negatively correlated with O_3. In Fig. 6 the diurnal cycles of NO_2 and NO are compared with O_3 diurnal cycle to provide further insight into the negative correlation between NOx and O_3. It can be clearly observed in Fig. 6 that while O_3 levels increase in the afternoon, the NO and NO_2 levels decrease and reach the minimal levels, probably due to photochemical dissociation of NO_2 in high solar radiation and enhanced dispersion process during afternoon. Furthermore, higher levels of NO and NO_2 are observed at about 08:00 - 09:00 am which are due to the peak traffic hours. Previously, several authors have used k-means clustering approach for ozone analysis in different countries around the world. For example, Munir (2013) used k-means to characterise O_3 diurnal cycle in the UK. Adame *et al.* (2012) applied k-means to surface O_3, NO_2 and SO_2 daily patterns in an industrial area in

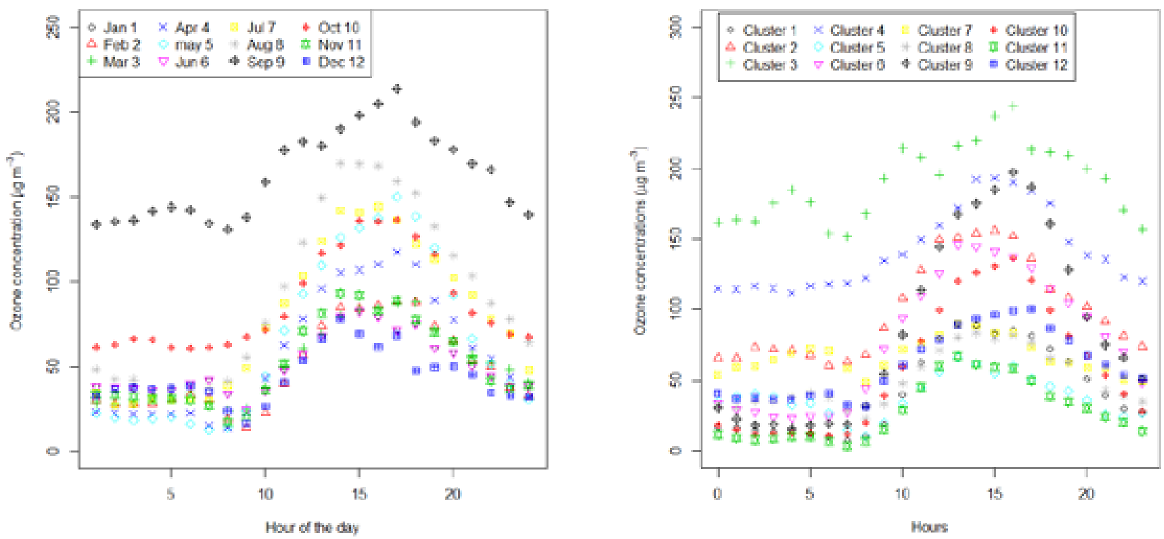

Figure 5. Diurnal cycles of O_3: during various months (January to December) (left); and using k-means clustering approach (right) at the PME station for year 2012.

Central-Southern Spain. They intended to find a set of representative daily cycles for each pollutant at different air quality regimes. They obtained five and four optimal cluster numbers for the daily patterns of O_3. Both Munir (2013) and Adame *et al*. (2012) found typical daily variations in O_3 concentrations. More recently, Austin *et al*. (2014) characterised O_3 trends in the USA with the help of meteorological parameters using k-means algorithm. They categorised days of observation based on the maximum daily temperature, standard deviation of daily temperature, mean daily ground level wind speed, mean daily water vapor pressure and mean daily sea-level barometric pressure. They found that O_3 trends were significantly different within the different weather groupings.

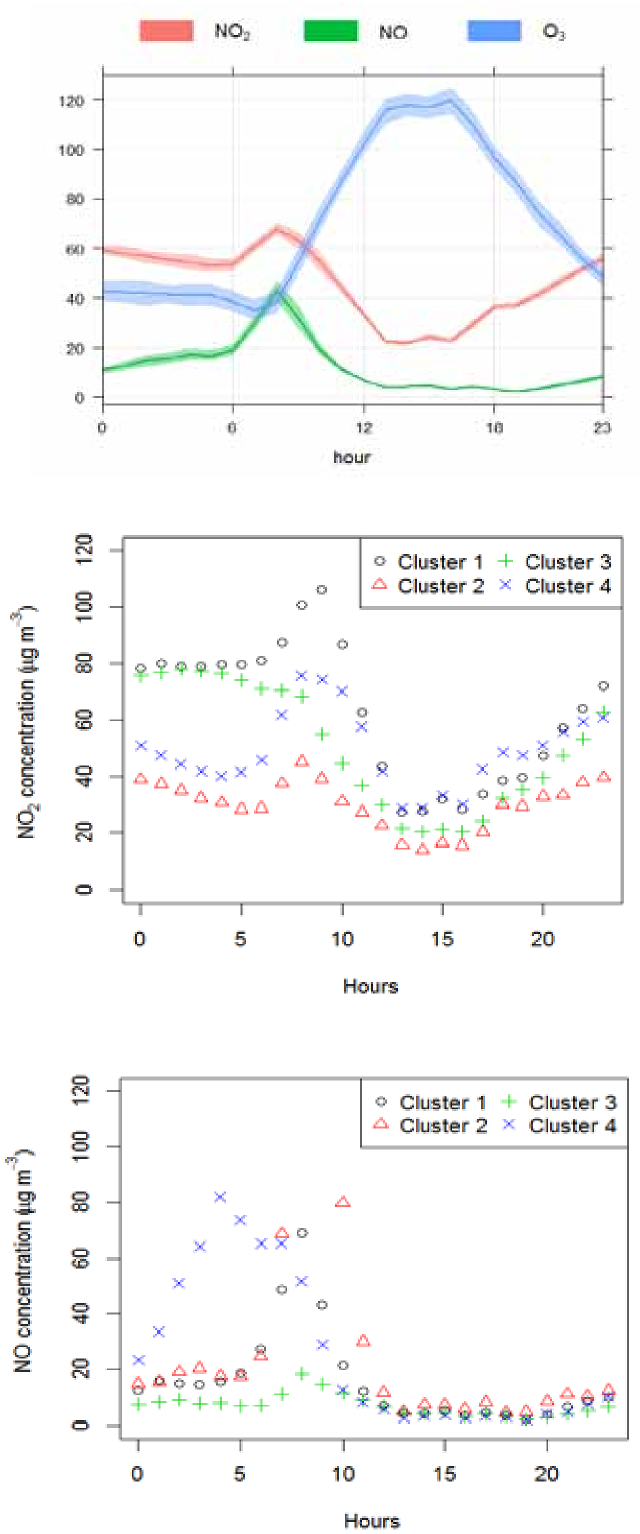

Figure 6. Comparison of the diurnal cycles of NO and NO_2 with O_3 concentrations ($\mu g/m^3$) (top), diurnal cycles of NO_2 (middle) and NO (bottom) using k-means clustering approach at the PME site for year 2012.

4. Conclusions

In this paper the reasons for temporal variations in O_3 concentrations during diurnal, weekly and seasonal cycles are discussed in the light of its precursors and meteorological parameters. It can be induced from the results of k-means algorithm that diurnal cycles of O_3 vary within the same month or season. In contrast, the same diurnal cycles can be observed in different months or seasons. These sorts of differences are not obvious when using only graphical presentations like time variation plots that only average concentrations over a period of time. Therefore, k-means clustering algorithm probably provides a useful tool for such analysis. The effect of temperature on O_3, particularly in the hottest months (June and July) requires further considerations to analyse the negative feedback of temperature on O_3 in the arid areas with the help of modelling approaches, which is part of the future work.

Acknowledgement

We are thankful to the Hajj Research Institute, Umm Al-Qura University Makkah, Saudi Arabia for sponsoring this study and for their continuous support for improving air quality conditions in the Holy Cities of Makkah and Madinah.

References

Adame JA, Notario A, Villanueva F, Albaladejo J. Application of cluster analysis to surface ozone, NO_2 and SO_2 daily patterns in an industrial area in Central-Southern Spain measured with a DOAS system. Science of the Total Environment 2012; 429(1): 281-91.

Andersson C, Langner J, Bergström R. Interannual variation and trends in air pollution over Europe due to climate variability during 1958-2001 simulated with a regional CTM coupled to the ERA40 reanalysis. Tellus 2007; 59(1): 77-98.

AQEG. Ozone in the UK, the fifth report produced by air quality expert group. Published by the Department for the Environment, Food and Rural Affairs 2009. DEFRA publication London, 2009AQEG.

Austin E, Zanobetti A, Coull B, Schwartz J, Gold DR, Koutrakis P. Ozone trends and their relationship to characteristic weather patterns. Journal of Exposure Science and Environmental Epidemiology 2014; online publication 9 July 2014.

Baur D, Saisana M, Schulze N. Modelling the effects of meteorological variables on ozone concentration - a quantile regression approach. Atmospheric Environment 2004; 38(28): 4689-99.

Carslaw D, Ropkins K. Openair - an R package for air quality data analysis. Environmental Modelling and Software 2012; 27(1): 52-61.

Coyle M, Smith RI, Stedman JR, Weston KJ, Fowler D. Quantifying the spatial distribution of surface ozone concentration in the UK. Atmospheric Environment 2002; 36(6): 1013-24.

Duenas C, Fernandez MC, Canete S, Carretero J, Liger E. Assessment of ozone variations and meteorological effects in an urban area in the Mediterranean Coast. Science of the Total Environment 2002; 299(1-3): 97-113.

Gardner MW, Dorling SR. Statistical surface ozone models: an improved methodology to account for non-linear behaviour. Atmospheric Environment 1999; 34(1): 21-34.

Habeebullah TM. An analysis of air pollution in makkah-a view point of source identification. Environment Asia 2013; 6(2): 11-17.

Hassan IA, Basahi JM, Ismail IM, Habeebullah TM. Spatial distribution and temporal variation in ambient ozone and its associated NOx in the atmosphere of Jeddah City, Saudi Arabia. Aerosol and Air Quality Research 2013; 13(6): 1712-22.

Jenkin ME. Trends in ozone concentration distributions in the UK since 1990: Local, regional and global influences. Atmospheric Environment 2008; 42(21): 5434-45.

Jenkin ME. Analysis of sources and partitioning of oxidant in the UK-Part1: The NOx-dependence of annual mean concentrations of nitrogen dioxide and ozone. Atmospheric Environment 2004; 38(30): 5117-29.

Jenkin ME, Davies TJ, Stedman JR. The origin and day of the week dependence of photochemical ozone episodes in the UK. Atmospheric Environment 2002; 36(6): 999-1012.

Kurokawa J, Ohara T, Uno I, Hayasaki M, Tanimoto H. Influence of meteorological variability on interannual variations of springtime boundary layer ozone over Japan during 1981-2005. Atmospheric Chemistry and Physics 2009; 9(2): 6287-304.

MacQueen JB. Some methods for classification and analysis of multivariate observations. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press 1967; 1: 281-97.

Munir S. Spatial-temporal analysis of traffic-related ground level ozone. Ph.D. thesis. University of Leeds, 2013. Available from: http://etheses.whiterose.ac.uk/5504/.

Munir S. Modelling the effect of warming climate on ground level ozone-a case study in Makkah. The 2nd Saudi International Conference of Environmental Technology (2nd SICET), Riyadh, Saudi Arabia on 10th to 12th of November 2014 (Invited talk).

Munir S, Chen H, Ropkins K. Characterising the temporal variations of ground level ozone and its relationship with traffic-related air pollutants in the UK: a quantile regression approach. International Journal of Sustainable Development and Planning 2012; 9(1): 603-13.

Munir S, Habeebullah TM, Seroji AR, Gabr SS, Mohammed AMF, Morsy EA. Quantifying temporal trends of atmospheric pollutants in Makkah (1997-2012). Atmospheric Environment 2013; 77: 647-55.

R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0, Available from: http://www.R-project.org/. R Version 2.14.

Steinera AL, Davisa AJ, Sillmana S, Robert C, Owena RC, Michalaka AM, Fiore AM. Observed suppression of ozone formation at extremely high temperatures due to chemical and biophysical feedbacks. Proceedings of the National Academy of Sciences of the United States of America 2010; 107: 19685-90.

Received 19 August 2014
Accepted 18 September 2014

Correspondence to
Assistant Professor Dr. Said Munir
The Custodian of the Hoy Mosques Institute for Hajj and Umrah Research,
Umm Al-Qura University,
Makkah,
Saudi Arabia
Tel: +9665 6734 9393
Fax: +966 2556 4955
E-mail: saidmunir@yahoo.co.uk