Modelling Particulate Matter in Makkah – A Quantile Regression Approach

نمذجة الأتربة العالقة في مكة المكرمة

إعداد:

Turki M. Habeebullah
Department of Environment and Health Research, The Custodian of the Two Holy Mosques Institute for Hajj and Umrah Research, Umm Al Qura University

Said Munir
Department of Environment and Health Research, The Custodian of the Two Holy Mosques Institute for Hajj and Umrah Research, Umm Al Qura University

Abstract

Atmospheric particulate matters, especially those with smaller sizes such as PM_{10} are associated with numerous adverse health and environmental impacts. PM_{10} concentrations are high in Makkah and exceed air quality standards. High PM_{10} levels might pose threat to human health, therefore for effective management it is vital to investigate and model the effect of different factors on PM_{10} concentrations in Makkah. In this paper for the first time in Makkah a quantile regression model (QRM) is developed using hourly PM_{10} concentrations (µg/m^3) as dependent variable and several air pollutants and meteorological variables as independent variables for year 2012. QRM addresses the problem of non-normal distributions of air quality data, and non-linearity in the association of PM_{10} with the covariates. All covariates show significant effect at least at one or more quantiles, however, wind speed, carbon monoxide (CO), nitric oxide (NO) and previous day PM_{10} (lag_PM_{10}) concentrations have significant effect at all quantiles and hence are considered the most important factors for controlling PM_{10} concentrations. Furthermore, CO has negative impact, whereas wind speed, NO and lag_PM_{10} have positive impact on PM_{10} concentrations. The strength, nature and direction of coefficients vary at different quantiles of the PM_{10} distribution. The performance of the model was assessed using several statistical metrics, including correlation coefficients (R, 0.82), factor of 2 (FACT2, 0.96), Root Mean Square Error (RMSE, 129), Normalised Mean Bias (NMB, 0.12) and Normalised Mean Gross Error (NMGE, 0.34). The values of these metrics and graphical presentations show that QRM performs better and explains significantly more variations in

PM_{10} concentrations than the multiple linear regression model (MLRM). To the best of our knowledge, this is the first study that uses a quantile regression approach for modelling PM_{10} levels in Makkah and probably elsewhere, and may help characterise and manage PM_{10} concentration in Makkah.

Objectives of the study

The aim of this research paper is to model PM_{10} concentration and investigate its association with other air pollutants and meteorological variables with the help of quantile regression model. The main objectives of the study are given below:

1. To analyse PM_{10} concentration and characterise its behaviour in Makkah;
2. To investigate the association of PM_{10} with other air pollutants;
3. to analyze PM10 concentration and characterise its behavior in relation to meteorological parameters in Makkah
4. To assess the performance of QRM for modelling PM_{10} in Makkah;
5. To compare the performance of QRM and MLRM for predicting PM_{10} in Makkah;
6. To emphasise the need for analysing the whole distribution of dependent variable, rather than only the central value (mean or median);
7. To help better manage PM_{10} and reduce its adverse impact on human health in Makkah.

Keywords: PM_{10}, air pollution, Makkah - Saudi Arabia, Quantile Regression Model.

Introduction

Air pollutants have negative impacts on human health, agricultural crops, ecosystem and building materials (e.g., Dockery et al., 1993; Burnett et al., 2000; WHO, 2004). Atmospheric particles aggravate chronic respiratory and cardiovascular diseases, alter host defence, damage lung tissues, lead to premature death, and possibly cause cancer (WHO, 2004; Hassan, 2006). Furthermore, particles have a range of important non-biological impacts, including soiling of man-made materials and buildings, reducing visibility and affecting heterogeneous atmospheric chemistry (Harrison, 2001). The adverse impacts of air pollutants are not limited to local areas where the pollutants are emitted and rather extend to regional and global levels in the form of acid rain and ground level ozone, which have transboundary impacts (e.g., AQEG, 2009; Hassan et al., 2013).

Makkah is one of the busiest cities in the world. Every year millions of people visit the city due to its religious importance. The high level of air pollutants is one of the growing concerns in Makkah, especially during the season of Hajj and Umrah as reported by several authors (e.g., Al-Jeelani, 2009; Othman et al., 2010; Seroji, 2011; Munir et al., 2013a; Munir et al., 2013b, Habeebullah, 2013). PM_{10} concentrations in Makkah exceed air quality standards set for the protection of human health. The reasons for the high particulate matter concentrations are most probably high volume of road traffic, construction work, resuspension of particles, windblown dust and sand particles, and geographical conditions (arid region) with hot temperature and low rainfall (Khodeir et al., 2012; Munir et al, 2013b). Furthermore, it is reported that the concentrations of PM_{10} in Makkah have increased during the last 15 years or so (Munir et al., 2013b).

PM_{10} levels are affected wind speed and direction, relative humidity, temperature, and rainfall (e.g., Baur et al., 2004; Elminir, 2005; Ordonez et al., 2005; Cheng et al., 2007; Beaver and Palazoglu, 2009; Pearce et al., 2011). Wind speed, turbulence level, air temperature, and precipitation affect the re-suspension of particles from the ground surface, their residence in the atmosphere, and the formation of secondary pollutants (Bhaskar and Mehta, 2011). Furthermore, other air pollutants, such as carbon monoxide (CO), sulphur dioxide (SO_2), and nitrogen oxide (NOx) can result in secondary aerosols formation, for example, SO_2 is oxidised in the atmosphere to form sulphuric acid (H_2SO_4), which can be neutralised by ammonia (NH_3) to form ammonium sulphate ($(NH_4)_2SO_4$). Similarly Nitrogen Dioxide (NO_2) is oxidised to nitric acid (HNO_3), which in turn can react with NH_3 to form ammonium nitrate (NH_4NO_3). Secondary sulphate (SO_4^{-2}) and nitrate (NO_3^-) particles are usually the dominant component of fine secondary particles (Harrison, 2001; WHO, 2003). Moreover, the interaction of these pollutants with each other and with PM_{10} can result in synergistic (positive interdependence) or antagonistic (negative interdependence) effects that can affect the adverse impact on human health and natural environment (WHO, 2003). How meteorology and other air pollutants affect the concentration of PM_{10} in an arid region like Makkah, where air quality data are limited, is not well characterised. Therefore, advanced modelling studies are required to analyse the effects of various controlling factors that can help in understanding and effective management of PM_{10} concentrations in Makkah and elsewhere.

Recently a generalised additive model (GAM) was developed to investigate the association of PM_{10} with various predictors in Makkah (Munir et al., 2013a), however GAM like multiple linear regression models focuses on the mean level of dependent variable (here PM_{10}) and fails to model the whole distribution, including extreme values, which are probably more important from public health point of view. In this paper, a quantile regression model (QRM) is employed to model PM_{10} concentrations. QRM is an advanced modelling approach that can help model the effect of independent variables on various quantiles of the dependent variable. QRM is also applicable to non-normal air quality distribution and can address the inherited non-linearities in the association between dependent and independent variables.

Methodology

Data source

This study uses a one year data measured at the Presidency of Meteorology and Environment (PME) monitoring station, situated near the Holy Mosque (Al-Haram) in Makkah, Saudi Arabia for the year 2012. The monitoring site and the air quality network in Makkah have been defined in Munir et al. (2013a and b) and are shown in. This study characterises PM_{10} concentration (µg/m^3) with the aid of several air pollutants (CO mg/m^3, SO_2 µg/m^3, NOx µg/m^3) and meteorological parameters (relative humidity (RH %), Temperature (T ^{o}C), wind speed (WS m/s), Wind Direction (WD degrees from the north) and atmospheric pressure (P) measured in hectopascal (hPa), which is equivalent to the conventional unit millibar (mbar). A summary of these parameters is presented in Table 1, showing minimum (min), 1st quartile (0.25 quantile), mean, median (0.5 quantile), 3rd quartile (0.75 quantile) and maximum levels of the given parameters. These parameters are continuously monitored at the PME monitoring site. Gaseous air pollutant levels can be expressed as mixing ratios [e.g., parts per million (ppm) or parts per billion (ppb)] or as concentrations (e.g., µg/m^3 or mg/m^3), however, PM_{10} are always expressed as concentration (e.g., µg/m^3). In this paper all pollutants are expressed as concentrations (μg/m^3 or mg/m^3) to be consistent in the use of units for both gaseous and non gaseous pollutants.

In order to make the collected data useful and to provide a sound scientific basis for comparison against air quality standards, public information or policy development, the data need to be accurate and reliable. Strict QA/QC (Quality Assurance and Quality Control) measures are taken to ensure the quality of data. This process makes sure that the data are (a) Genuinely representative of atmospheric concentrations in the areas under investigation; (b) Representative over the period of measurement. Data capture is greater than 90 % for all parameters, except SO_2 where 88 % data were present.

It is shown in Figure 2 that PM_{10} concentrations and independent variables are not normally distributed. The histograms are right (positive) skewed. This has been reported previously by several authors (Duenas et al., 2002; Munir et al., 2011) elsewhere that air pollutants and meteorological variables are not normally distributed. The majority of classical statistical tests are based on the assumption that the data to which the tests are applied should exhibit a normal distribution (i.e. bell shape, symmetrical and with a common mean and median). If the parametric tests are applied to non-normal data, they can result in biased or even erroneous results (Reiman et al., 2008). Therefore, before applying a classical test it is vital to check data distributions and if the data are non-normally distributed, robust and non-parametric methods should be applied that are not based on such assumptions.

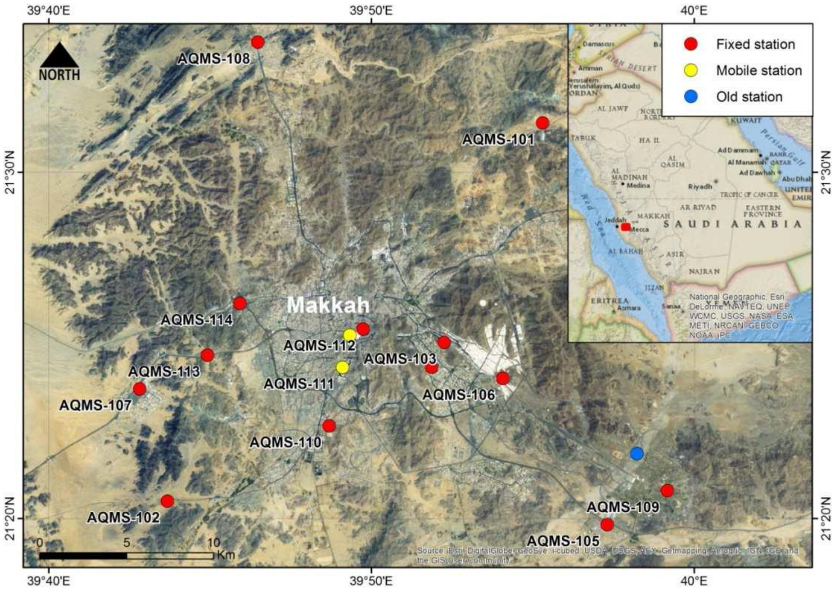


Figure 1. Map of the air quality monitoring sites in Makkah (Munir et al., 2013b).

Table1. Showing a summary of the parameters used in this study measured at the PME monitoring station near the Holy Mosque in Makkah, Saudi Arabia for the year 2012.

Pollutant	Min	1^{st} quartile	Mean	Median	3^{rd} quartile	Maximum	%data capture
[1]CO (mg/m^3)	0	0.79	0.98	1.12	1.27	6.87	95
SO_2 ($\mu g/m^3$)	0	5	8	11	15	125	88
NO_2 ($\mu g/m^3$)	0	27	42	46	61	223	99
NOx ($\mu g/m^3$)	0	21	33	42	52	367	99
NO ($\mu g/m^3$)	0	2	5	12	13	299	99
PM_{10}($\mu g/m^3$)	0	79	124	180	199	5761	93
P (hPa)	649	971	975	974	978	984	100
RH (%)	4	18	31	33	45	86	100
T (^{0}C)	16	27	32	32	36	46	100
WS (m/s)	0	1	1	1	2	6	100
WD (degrees)	1	185	285	243	333	360	100

[1]In the table SO_2 stands for sulphur dioxide, CO for carbon monoxide, NO for nitric oxide,

NO_2 for nitrogen dioxide, NOx for nitrogen oxides, O_3 for ozone, PM_{10} for particles with aero dynamic diameter of 10 um or less, WS for wind speed, WD for wind direction, T for temperature, RF for rainfall, RH for relative humidity, and P for atmospheric pressure.

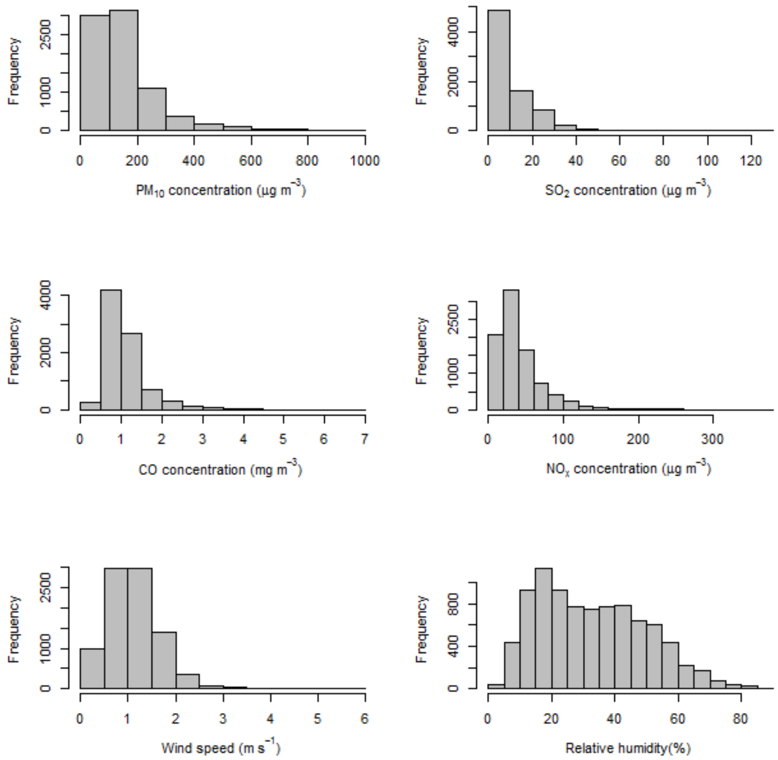


Figure 2. Histograms showing the frequency distributions of mean hourly data of PM10, SO_2, CO, NOx, wind speed, and relative humidity at the PME (Presidency of Meteorology and Environment), near the Holy Mosque in Makkah, Saudi Arabia for the year 2012.

General Statistics

Statistical Software R programming language (R Development Core Team, 2012) and two packages Quantreg, version 4.9.1 (Koenker, 2012) and openair version 2.13.2 (Carslaw and Ropkins, 2012) are used for running QRM, performing other statistical analysis and making graphs. Graphical presentations (e.g., histograms, polar plot and scatter diagram) are also used to present the outputs of the analysis.

Quantile Regression Model (QRM)

In this paper QRM model is employed to analyse the effect of covariates (e.g., meteorological parameters and other air pollutants such as NOx, CO, SO_2) on PM_{10} concentrations. QRM allows the covariates to have different contribution at different quantiles of the dependent variable distribution (here PM_{10}) and is robust (insensitive) to departures from normality and to skewed tails. Air pollutant data are not normally distributed as reported by several authors (e.g., Duenas et al., 2002; Munir et al., 2011) and is also demonstrated in Figure 1. Furthermore, air pollutants exhibit nonlinear association with its predictors (e.g., Gardner and Dorling, 2000; Baur et al., 2004). This means that the contributions of the explanatory variables (e.g., meteorological variables) to PM_{10} concentration vary significantly at different levels. This suggests that statistical models should have the capability to address the linearity and normality issues when applying to analyse PM_{10} data. QRM is capable of addressing these issues. Readers are referred to Koenker (2005) and Hao and Naiman (2007) for details on QRM; and to Baur et al. (2004) and Munir et al. (2012) for the applicability of QRM to ground level ozone concentrations. Baur et al. (2004) modelled the impact of meteorology on ozone concentration in Athens, whereas Munir et al. (2012) modelled the effect of road traffic on ozone concentrations in the UK.

Using hourly mean PM_{10} concentrations as a dependent (modelled or response) variable, and several meteorological parameters (T, RH, P, WS, and WD) and air pollutants (CO, NO, NO_2, SO_2 and lag_PM_{10}) as independent variables, a QRM is developed and compared with Multiple Linear Regression Model (MLRM). These covariates are important for modelling PM_{10} concentrations and control a significant proportion of PM_{10} variations as previously shown by Munir et al. (2013a). MLRM specifies the conditional mean function, whereas QRM specifies the conditional quantile function. MLRM and QRM are shown below in Equations (1) and (2), respectively (Hao and Naiman, 2007).

$$PM_{10} = \beta_o + \beta_1 P + \beta_2 RH + \beta_3 T + \beta_4 WS + \beta_5 WD + \beta_6 CO + \beta_7\ SO_2 + \beta_8\ NO + \beta_9\ NO_2 + \beta_{10}\ lag\_PM_{10} + \varepsilon_i \ ...... (1)$$

$$PM_{10} = \beta_o^{(p)} + \beta_1^{(p)}P + \beta_2^{(p)}RH + \beta_3^{(p)}T + \beta_4^{(p)}WS + \beta_5^{(p)}WD + \beta_6^{(p)}CO + \beta_7^{(p)}SO_2 + \beta_8^{(p)}NO + \beta_9^{(p)}NO_2 + \beta_{10}^{(p)}lag\_PM_{10} + \varepsilon_i \text{ ...... (2)}$$

In Equations (1) and (2) β_o represents the intercept, β_1 to β_{10} the slopes (gradients) of the covariates and ε_i the error term. The (p) shows the *p*th quantile and its value lies between 0 and 1. Equation (1) gives one coefficient for each variable, on the other hand equation (2) can have numerous quantiles and will require a separate equation for each quantile and therefore will produce numerous coefficients for each variable. This study adopts 11 quantiles (0.05, 0.1 – 0.9, 0.95) and therefore 11 equations will generate the same number of quantile regression coefficients for each covariate. Several metrics are calculated to assess the model performance. These metrics are: Root Mean Square Error (RMSE), Normalised Mean Gross Error (NMGE), Correlation coefficient (R), Normalised Mean Bias (NMB), and Factor of 2 (FAC2). For more details on these metrics, their definition and their mathematical formulae see Carslaw (2011) and Derwent et al. (2010).

QRM makes several predictions, one for each quantile and therefore the metrics used for assessing the model performance can be calculated for each quantile. The metrics are called local metrics, e.g., local goodness of fit, local MRSE and local FB etc. The local metrics cannot be compared with the metrics estimated for MLRM, as they have different nature (Baur et al., 2004). Therefore, global metrics need to be estimated for QRM to take account of all quantiles and make them comparable with MLRM. To estimate global metrics for QRM, this study adopts the amalgamated quantile regression model (AQRM) technique suggested by Baur et al. (2004). However, Baur et al. (2004) have used only coefficient of determination (R^2) value for assessing the model performance, whereas this paper extends this concept further to other metrics (NMB, NMGE, RMSE, FAC2). The first step is to run QRM and determine quantile regression coefficients for all the quantiles used in the model. QRM will normally give numerous predictions according to the number of quantiles. To turn those into one global prediction, the dataset is divided into the same number of subsets as the number of quantiles and then the model for that respective quantile is used to predict PM_{10} concentration. The predicted PM_{10} concentration for these quantiles is then re-integrated in such a way that it corresponds to the observed concentrations in the exact order. This gives a global prediction (prediction taking into account all quantiles), which is compared with observed concentration to calculate various metrics for assessing the performance of the model using various metrics according to the formulae given by Carslaw (2011) and Derwent et al. (2010).

Results and Discussions

The outputs of QRM (Equation 2) and MLRM (Equation 1) are depicted in Figure 3, which shows the effect of various covariates on PM_{10} concentration. The quantiles used in this study are shown on x-axis and their respective coefficients (slopes) are shown on y-axis. The dashed-dotted black line represents the coefficients of QRM, the solid red line represents the coefficient of MLRM and the solid black is the zero line. When any of the confidence intervals overlaps with zero line, it shows non-significant effect and vice versa. Understandably negative coefficients show negative effect, whereas positive coefficients show positive effect of the independent variables on PM_{10} concentrations.

The first panel in Figure 3 shows the intercept of the model. The intercepts are within the range of +100 and -113 for quantile 0.9 and 0.8, respectively, except quantile 0.95 which has higher intercept. Positive coefficients show positive effect whereas negative coefficients show negative effect of the covariates on the PM_{10} concentrations. The effect of atmospheric pressure (Figure 3, top-middle panel) is significant only at quantile 0.95 and for the rest of the quantiles the confidence intervals overlap with zero line, showing non-significant effect. Significant negative effect at quantile 0.95 may be due to the fact that high PM_{10} concentration in Saudi Arabia is linked with high wind speed which in turn is associated with low pressure. This means high PM_{10} concentration is linked with low atmospheric pressure. It is worth mentioning here that quantile 0.95 is related to high PM_{10} concentration and not with high atmospheric pressure, however the negative coefficients signify negative association of pressure with PM_{10} concentration. Relative humidity shows significant negative mean (average) effect, which is significantly different from the effect at various quantiles. Furthermore, the negative effect of relative humidity is significant at quantiles 0.05 to 0.3 and negative but non-significant at higher quantiles. As reported previously by Munir et al. (2013a) that high relative humidity is generally linked with nighttime hours when dust concentration is generally low and therefore shows negative correlation with PM_{10} concentrations. Furthermore, high relative humidity might be related with precipitations which wash out the atmospheric particles. Duenas et al. (2002) has reported that relative humidity plays an important role in the overall reactivity of the atmospheric system, either by affecting chain termination reactions or in the production of wet aerosols, which in turn affect the flux of ultraviolet radiation. Furthermore, relative humidity is also considered to be a limiting factor in the disposition of NO_2 because high percentages of humidity favour the reaction of the NO_2 with salt particles, e.g., sodium chloride. Barmpadimos et al. (2011) have reported that the relationship between PM_{10} and relative humidity is not the same for different monitoring sites. They have shown that the nature of relationship between relative humidity and PM_{10} changed at various monitoring sites and also at different levels of the relative humidity, e.g., the association was positive at low relative humidity (< 60%) and negative at high relative humidity (> 60%).

The effect of temperature on PM_{10} concentration is insignificant at extreme values (top and bottom 10%) and significant at the middle quantiles (0.2 to 0.8), where the effect is positive. High temperature can results in enhanced resuspension of soil and road dust, and formation of secondary aerosol, hence a temperature increase from 10 to 35 ^{o}C increases PM_{10} concentration by a factor of 4 in warm days during summer (Barmpadimos et al., 2011). High levels of PM_{10} (extreme levels) in Makkah is mostly caused by sand storms and construction activities near the monitoring site (Munir et al., 2013b), which are not dependent on temperature as much as on wind speed and direction, therefore probably temperature show non-significant effect. The mean effect estimated by MLRM is negative, where the regression coefficient is about -2. Mean can be biased by outliers and therefore the results of MLRM can be confusing sometimes. This probably shows that for air quality analysis more robust metrics (e.g., median or other quantiles) should be used, which are not affected by extreme values in the concentrations. When temperature was used as the only model input, the effect became positive. This might mean that the effect of temperature changes when other inputs are added to the model, probably due to interaction of various input variables. The effect of wind speed is positive and significant at all quantiles. The effect of wind speed is much stronger: the coefficient at quantile 0.95 is about 120. The effect gradually increases as PM_{10} concentration increases; however, the rate of increase is greater at higher quantiles. The stronger effect of wind speed at higher PM_{10} concentration is expected as high wind speed blows sand and dust particles from the barren desserts around the Makkah city causing sand-and-dust storms. The effect of wind direction is positive at lower quantiles until quantile 0.7 and becomes negative at higher quantiles. Because of the circular nature of wind direction, its effect is more complicated and is further investigated with the help of polar plots (Figure 4). The plots are constructed by averaging pollutant concentration by wind speed categories (0–1 m/s, 1–2 m/s, etc.) as well as wind direction (0–10, 10–20, etc.). In polar plots the levels of PM_{10} concentration is shown as a continuous surface, which is calculated through using Generalized Additive Models *s*moothing techniques (Carslaw and Ropkins, 2012). It can be observed in Figure 4 that highest PM_{10} concentration is related with high wind speed (5 – 6 m/s) from the southeast direction. In addition at a wind speed about 3 m/s relative high PM_{10} concentration is shown in the west, northwest and east direction. Mostly low PM_{10} concentration can be observed at low wind speed (< 2 m/s) from all direction. Further investigation of the local area revealed that there was a large construction work going on near the Holy Mosque in the west to northwest direction. There are some barriers between the monitoring site and construction location; however it seems like when westerly wind blows at a speed greater than 2 m/s, the dusts manage to reach the monitoring site. On the eastern side, there is a busy road (Masjid Al-Haram road) and a couple of bus stations, which probably contribute to the monitored concentration.

CO shows negative effect on PM_{10} and the strength of coefficients (in absolute terms) increase as PM_{10} concentration increases. The effect of CO is significant at all quantiles and coefficients range from -8 to -47 at quantile 0.05 and 0.95, respectively. Mean regression coefficient was -60, which is stronger than the quantile coefficients; however it is not significantly greater than the coefficients of quantile 0.9 and 0.95. The effect of SO_2 is negative and significant at most of the quantiles, except at quantile 0.05, 0.8 and 0.9. Mean regression coefficient is about -2 and is significantly greater than the quantile regression coefficients. The positive effect of NO_2 is significant at quantile 0.05 to 0.6, whereas at higher quantiles (0.7 to 0.95) the effect is insignificant. On the other hand the effect of NO is positive and significant at all quantiles. Furthermore, the strength of coefficients gradually increases from quantile 0.05 to 0.95, in contrast to NO_2, where the strength of coefficients shows the opposite pattern. The effect of lag_PM_{10} is also positive, as expected and the effect becomes stronger as the concentration of PM_{10} increases. Fine and extra fine particles stay in the atmosphere for long time and contribute positively to the measured concentration hours or even days later (Munir et al., 2013a) probably that is why lag_PM_{10} demonstrate positive effect.

It can be observed in Figure 3 that the effect of independent variables on PM_{10} concentration is not linear and changes as the concentration of PM_{10} changes. Sometimes only the strength of coefficients changes and the nature (positive or negative) remains unchanged as in the case of wind speed, CO, NO and lag_PM_{10}, whereas other times both strength and nature of the coefficients change as in the case of atmospheric pressure, temperature and wind direction. It is shown that independent variables can have significant effect at some quantiles and insignificant at other quantiles (e.g., pressure, relative humidity, temperature, wind direction, SO_2 and NO_2), however, wind speed, CO, NO and lag_PM_{10} have significant effects at all quantiles. The insignificant effect is mostly related with high quantiles, for instance in the case of relative humidity, temperature, NO_2 and SO_2, however, temperature, pressure and SO_2 also show insignificant effect at lower quantiles. This sort of relationship usually remains hidden when applying linear models, e.g. MLRM, which assume linear association between dependent and independent variables..

Other air pollutants (e.g., CO and SO_2) would be expected to show positive correlation with PM_{10} because they have the same sources of emissions (e.g., road traffic in urban areas or at roadside locations) and also can add to secondary air pollutant formation, for example the conversion of SO_2 and NOx to sulphate (SO_4^{-2}) and nitrate (NO_3^-) ions, respectively. However, here the association is predominantly negative, particularly the association of CO and SO_2 with PM_{10}. To investigate this further, scatter plots of CO, SO_2 and NOx against PM_{10} are shown in Figure 5, where the number of data points are colour coded to show where most of the data points lie (left to right and top to bottom, panel 1- 3). Figure 5 clearly shows (left to right and top to bottom, panel 4-6) two different patterns. The red colour shows high PM_{10} concentrations associated with low concentrations of other pollutants (e.g., CO). The blue colour indicates a different pattern, i.e. as the concentrations of other pollutants increase, there seems to be little variation in PM_{10} concentrations. As mentioned above wind speed, probably plays the dominant role in the negative association of PM_{10} with other air pollutants. High wind speed, blowing sand and dust particles enhances the concentration of PM_{10} and disperses locally emitted gaseous pollutants. Thus episodes of high PM_{10} are associated with low levels of other pollutants and vice versa, which probably explains the negative effect of CO and SO_2 on PM_{10} concentration (Figure 5).

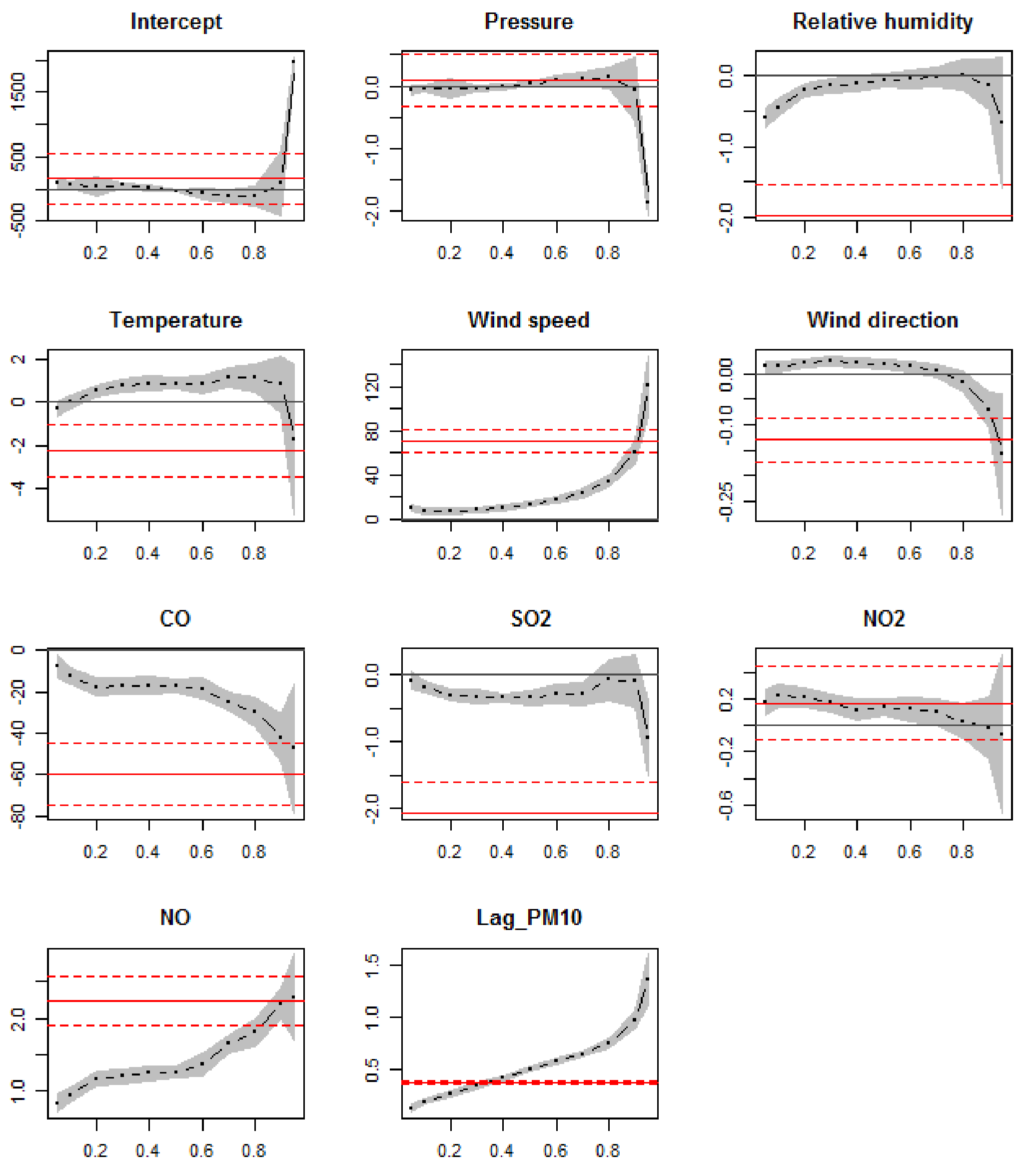


Figure 3. The outputs of quantile regression model showing the effect of atmospheric pressure (hPa), relative humidity (%), temperature (°C), wind speed (m/s), wind direction (degrees from the north), carbon monoxide (CO mg/m^3), sulphur dioxide (SO_2 µg/m^3), nitrogen dioxide (NO_2 µg/m^3), nitric oxide (NO µg/m^3) and lag_PM_{10} (previous day PM_{10} concentrations µg/m^3) on PM_{10} concentration (µg/m^3). Quantile regression coefficients (dashed dotted dark line) and ordinary least square regression coefficients (solid red line) are presented with their 95% confidence interval. Various quantiles are shown on x-axis and their respective coefficients on y-axis.

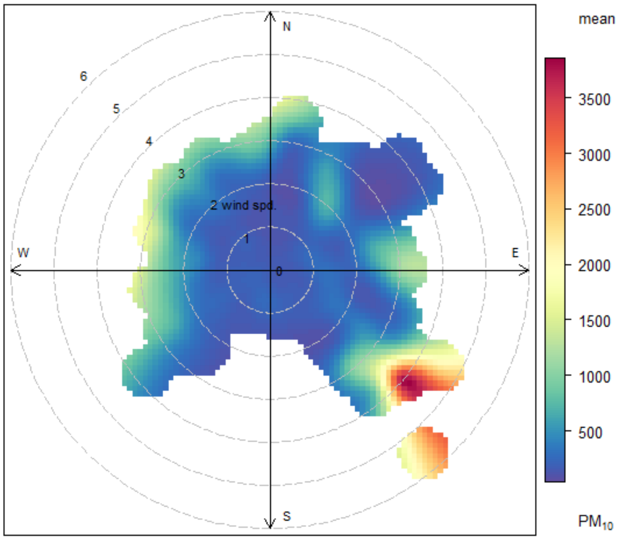
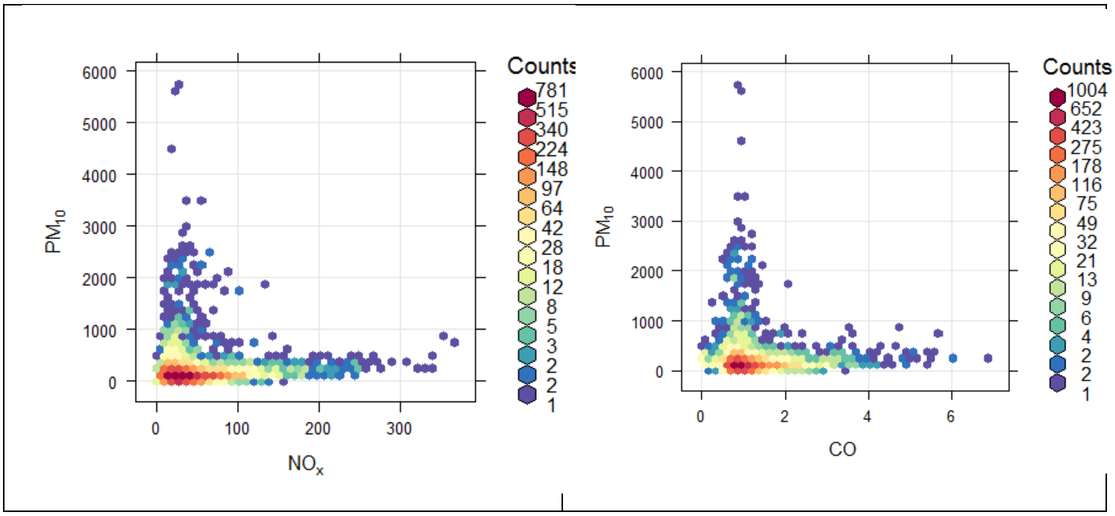


Figure 4. Polar plot of PM_{10} concentration (µg/m^3) near the Holy Mosque, Makkah, colour coded by PM_{10} concentrations for 2012.



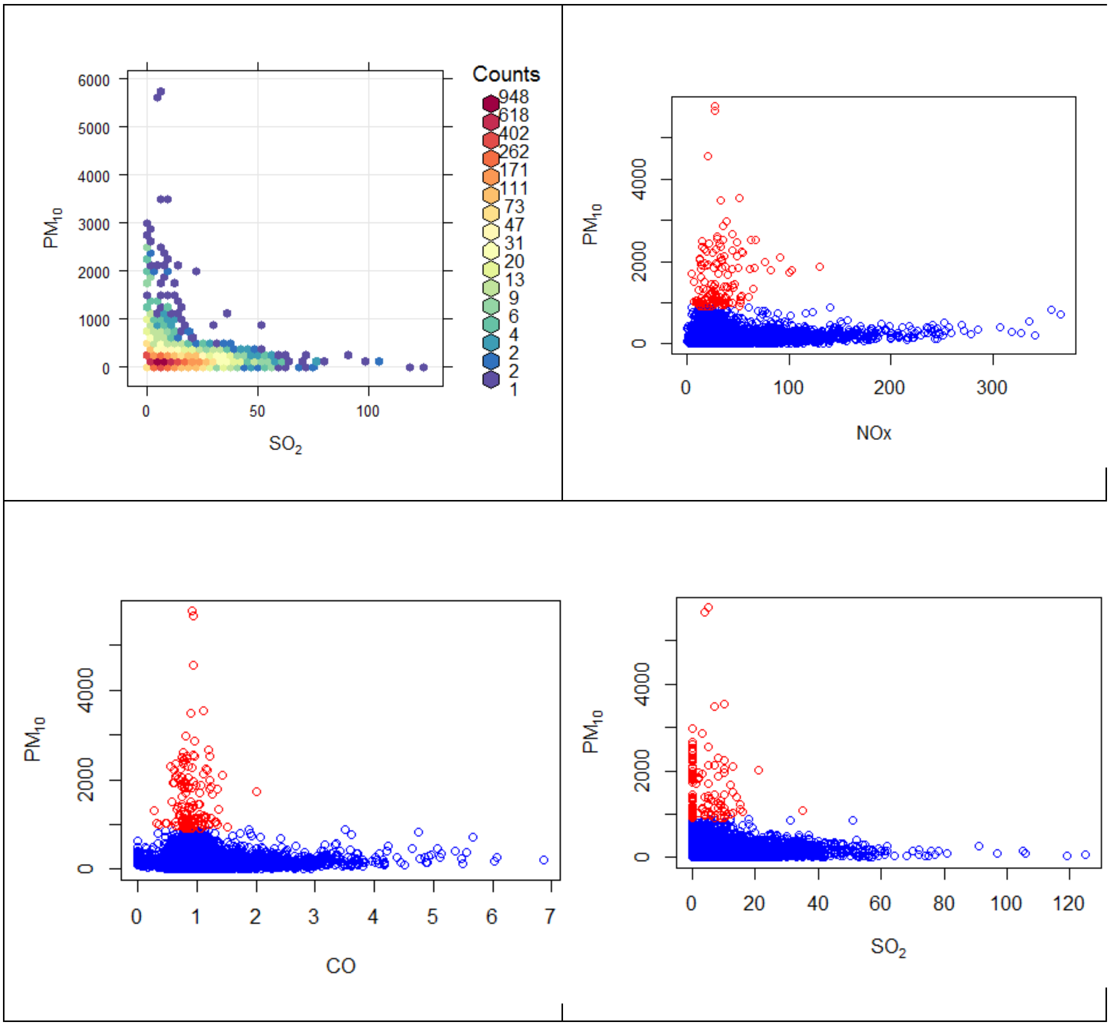


Figure 5. Scatter plots of hourly PM_{10} concentrations ($\mu g/m^3$) versus NOx ($\mu g/m^3$), CO (mg/m^3) and SO_2 ($\mu g/m^3$) concentrations measured at PME monitoring stations near the Holy Mosque in Makkah, Saudi Arabia, 2012. ‘Counts’ shows the number of data points, meant to present where most of the data points lie. In the lower panels the red and blue colour indicates different patterns in the association of PM_{10} and other pollutants.

To assess the QRM performance Hao and Naiman (2007) have suggested the estimation of local coefficient of determination ($R^1\tau$), which are calculated for each quantile considered in the model. However, $R^1\tau$ cannot be compared with normal coefficient of determination (R^2) and therefore do not aid much in assessing the model performance. In this study the approach recommended by Baur et al. (2004) for calculating R^2 is adopted and extended further by calculating several other metrics as suggested by Carslaw (2011) and Derwent et al. (2010). An air quality model is considered acceptable if more than half of the predicted values are within a factor of 2 of the observed concentration and faulty if not. Furthermore, it is recommended that air quality models are considered acceptable if NMB values lie within the range between -0.2 and +0.2 and faulty otherwise (Derwent et al., 2010). Table 2 shows that both of these metrics for both QRM and MLRM are within in the recommended range, and hence the performance of the models is acceptable. In addition, the performance of the QRM is better than that of MLRM, for instance FAC2 and correlation coefficients for QRM and MLRM are 0.96, 0.82 and 0.82, 0.39, respectively.

Figure 6 compares observed and predicted PM_{10} concentrations of both QRM and MLRM with the help of a scatter plot, which is very useful for model evaluation (Carslaw, 2011). In the scatter plot it is much easier to see where the data lie and to get a feeling about bias, etc. Relatively more points lie below the 1:1 line (middle line in Figure 5) in case of MLRM and there seems to be a slight negative bias (under prediction); whereas more points lie above the 1:1 line in case of QRM, showing slight positive bias (over prediction). Particularly at high concentration of PM_{10} MLRM fails to perform and under predicts PM_{10} concentration. The dashed lines show the within factor of two (FAC2) region and it is perhaps worth noting that majority of points lie well within this region.

Table 2. Statistical metrics for assessing the performance of the model calculated for the testing dataset June 2012 at PME monitoring station in Makkah.

Metric	[1]QRM	MLRM
FAC2	0.96	0.82
MB	25.71	-14.35
MGE	69.66	104.58
NMB	0.12	-0.06
NMGE	0.34	0.43
RMSE	129.06	204.34
R	0.82	0.39

[1]QRM stands for quantile regression model; MLRM for multiple regression model; FAC2 for Factor of Two; MB for Mean Bias; RMSE for Root Mean Square Error; MGE for Mean Gross Error; NMGE for Normalised Mean Gross Error; R for correlation coefficient; and NMB for Normalised Mean Bias. For definitions and calculation methods of these metrics see Derwent et al. (2010) and Carslaw (2011).

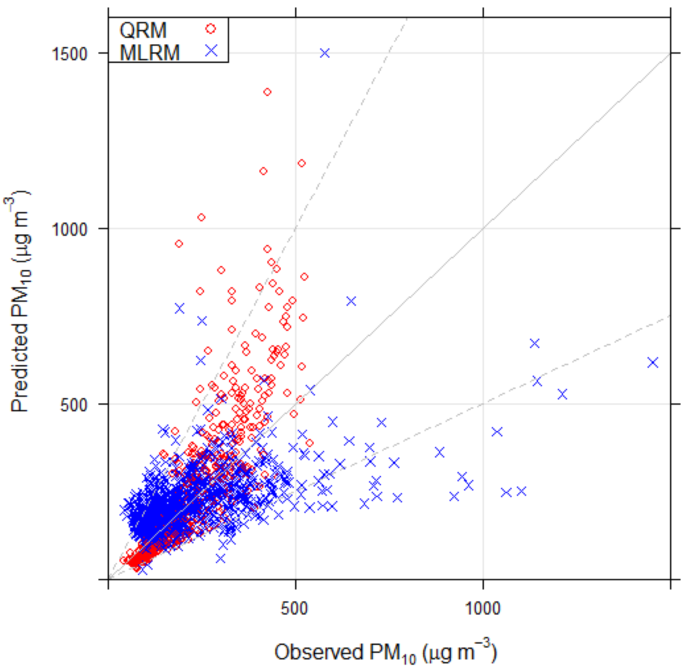


Figure 6. Comparison of observed and predicted PM_{10} concentrations (µg/m^3) based on the testing dataset for June 2012. The middle solid line is 1:1, and the above and below dashed lines are 0.5:1 and 2:1, respectively. So, the area between the two dashed lines is the factor of two (FAC2) regions.

4. Conclusions and Recommendation

This study employs a quantile regression model to characterise the effect of several air pollutants and meteorological variables on PM_{10} concentration in Makkah, Saudi Arabia. QRM characterises the effect of covariates at various quantiles, in contrast to the traditional approaches which investigate the effect of independent variables on the mean of the dependent variable (here PM_{10}). The effect of the independent variables (pressure, relative humidity, temperature, wind speed, wind direction, CO, SO_2, NO, NO_2, and lag_PM_{10}) was significant at least at one or more quantiles of the PM_{10} concentrations. However, the effect of wind speed, CO, NO and lag_PM_{10} was significant at all quantiles and hence seems to be controlling most of the variations in PM_{10} concentrations. Scatter plots and polar plots were employed to provide further insight into the association of these variables with PM_{10} concentration. The model performance is assessed by calculating several statistical metrics, including R (0.82), MRSE (129), FAC2 (0.96), MB (25.71), NMB (0.12), MGE (69.66) and NMGE (0.34) and graphical presentation. The values of these metrics show satisfactory performance of the models, especially that of QRM, which analyses the whole PM_{10} distribution and is therefore recommended for modelling PM_{10} in Makkah. As a result of this study the following recommendation can be given:

1. QRM analyses the whole PM_{10} distribution and outperforms MLRM and is therefore preferred over the traditional linear approach;
2. Multiple linear regression model fails to capture variations in PM_{10} concentrations and is not recommended for modelling PM_{10} in Makkah;
3. This study uses data from only one monitoring station, therefore it is recommended to collect data at several sites in Makkah to provide a full picture of the spatial variations of PM_{10} in Makkah.
4. No air pollutant emission sources (e.g. road traffic) data were available, which could further improve the performance of the model. Further work is required to quantify the emission of PM_{10} from various sources and analyse their impact on observed concentrations;
5. In addition to statistical models, dispersion model should be run to characterise different air pollutants and prepare an effective air quality management plan in Makkah.

Acknowledgement

This work is funded by Hajj Research Institute (HRI), Umm Al-Qura University Makkah, Saudi Arabia. The research was conducted for the improvement of air quality in Makkah and Madinah. Thanks are also extended to PME for providing us the data and to all staff of the HRI, particularly to the staff of the Department of Environment and Health Research for their support.

References

Al-Jeelani, H.A., 2009. Evaluation of Air Quality in the Holy Makkah during Hajj Season 1425 H. Journal of Applied Sciences Research 5: 115-121.

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

Baur, D., Saisana, M., Schulze, N., 2004. Modelling the effects of meteorological variables on ozone concentration-a quantile regression approach. Atmospheric Environment 38: 4689 – 4699.

Beaver, S., Palazoglu, A., 2009. Influence of synoptic and mesoscale meteorology onozone pollution potential for San Joaquin Valley of California. AtmosphericEnvironment 43 (10): 1779-1788.

Bhaskar, B.V., Mehta, V.M., 2011. Atmospheric Particulate Pollutants and their Relationship with Meteorology in Ahmedabad. Aerosol and Air Quality Research, 10: 301–315.

Burnett, R.T., Brook, J., Dann, T., Delocla, C., Philips, O., Cakmak, S., Vincent, R., Goldberg, M.S., Krewski, D., 2000. Association between particulate- and gas-phase components of urban air pollution and daily mortality in eight Canadian cities. Inhalation Toxicology 12 (Suppl. 4), 15–39.

Carslaw, D., 2011. Defra regional and transboundary model evaluation analysis – Phase 1. Version: 15th April 2011.

Carslaw, D., and Ropkins, K., 2012. Openair - An R package for air quality data analysis. Environmental Modelling & Software 27: 52-61.

Cheng, C.S.Q., Campbell, M., Li, Q., Li, G., Auld, H. Day, N., Pengelly, D., Gingrich, S., Yap, D., 2007. A synoptic climatological approach toassess climatic impact on air quality in South-central Canada. Part I: historicalanalysis. Water Air and Soil Pollution 182 (1-4): 131-148.

Derwent, D., Fraser, A., Abbott, J., Jenkin, M., Willis, P., and Murrells, T., 2010. Evaluating the performance of air quality models. Issue 3/June 2010.7, 81.

Dockery, D.W., Pope, C.A., Xu, X., Spengler, J.D., Ware, J.H., Fay, M.E., Ferris, B.G., Speizer, F.E., 1993. An association between air pollution and mortality in six US cities. New England Journal of Medicine 329, 1753–1759.

Duenas, C., Fernandez, M.C., Canete, S., Carretero, J., Liger, E., 2002. Assessment ofozone variations and meteorological effects in an urban area in the MediterraneanCoast. The Science of the Total Environment 299, 97-113.

Elminir, H.K., 2005. Dependence of urban air pollutants on meteorology. Science ofthe Total Environment 350 (1-3), 225-237.

Gardner, M.W., Dorling, S.R., 2000. Statistical surface ozone models: an improved methodology to account for non-linear behaviour. Atmospheric Environment 34, 21 – 34.

Habeebullah, T.M., 2013. Health Impacts of PM_{10} Using AirQ2.2.3 Model in Makkah, Journal of Basic and Applied Sciences, *9*, 259-268.

Hao, L., Naiman, D.Q., 2007. Quantile regression : Series-Quantitative applications in the social sciences, Sage Publications, 2007 (Series NO. 07- 149).

Harrison, R.M., 2001. In Pollution causes, effects, and control. Fourth Edition, Royal Society of Chemistry, R.M. Harrison (Ed.).

Hassan, I.A., Basahi, J.M., Ismail, I., Habeebullah, T.M., 2013. Spatial Distribution and Temporal Variation in Ambient Ozone and Its Associated NOx in the Atmosphere of Jeddah City, Saudi Arabia. Aerosol and Air Quality 13: 1712 – 1722. doi: 10.4209/aaqr.2013.01.0007

Hassan, S.K.M., 2006. Atmospheric polycyclic aromatic hydrocarbons and some heavy metals in suspended particulate matter in urban, industrial and residential areas in Greater Cairo.Ph.D. Thesis, Chemistry Department, Faculty of Science, Cairo University, Egypt.

Khodeir, M., Shamy, M., Alghamdi, M., Zhong, M., Sun, H., Costa, M., Chen, L.C., Maciejcczyk, P.M., 2012. Source Apportionment and Elemental Composition of PM2.5 and PM_{10} in Jeddah City, Saudi Arabia. Atmospheric Pollution Research 3: 331–340.

Koenker, R., 2005. Quantile regression, Economic Society MonographsNo. 38. Cambridge University press. ISBN 0-521-608227-9.

Koenker, R., 2012. quantreg: Quantile Regression. R package version 4.91 (http://CRAN.R-project.org/package=quantreg).

Munir, S., Chen, H., Ropkins, K., 2011. Non-parametric nature of tropospheric ozone and its dependence on nitrogen oxides: a view point of vehicular emission. In: Brebbia, C.A., Longhurst, J.W.S., Popov, V. (Eds.), 2011a. Air Pollution XIX, vol.147. WIT Press, ISBN 978-1-84564-528-1, pp. 93-104.

Munir, S., Chen, H., Ropkins, K., 2012. Modelling the impact of road traffic on ground level ozone concentration using a quantile regression approach, Atmospheric Environment, Atmospheric Environment 60: 283-291.

Munir, S., Habeebullah, T.M., Seroji, A.R., Morsy, E.A., Mohammed, A.M.F., Saud, W.A., Esawee, A.L., Awad, A.H., 2013a. Modelling Particulate Matter Concentrations in Makkah, Applying a Statistical Modelling Approach, Aerosols Air Quality Research 13: 901-910.

Munir, S., Habeebullah, T.M., Seroji, A.R., Gabr, S.S., Mohammed, A.M.F., Morsy, E.A., 2013b. Quantifying temporal trends of atmospheric pollutants in Makkah (1997 – 2012). Atmospheric Environment 77: 647 - 655.

Ordonez, C., Mathis, H., et al., 2005. Changes of daily surface ozone maxima in Switzerland in all seasons from 1992 to 2002 and discussion of summer 2003. Atmospheric Chemistry and Physics 5, 1187-1203.

Othman, N., Mat-Jafri, M.Z., San, L.H., 2010. Estimating Particulate Matter Concentration over Arid Region Using Satellite Remote Sensing: A Case Study in Makkah, Saudi Arabia. Modern Applied Science 4: 11 - 20.

Pearce, J.L., Beringer, J., Nicholls, N., Hyndman, R.J., Tapper., N.J., 2011. Quantifying the influence of local meteorology on air quality using generalized additive models. Atmospheric Environment 45: 1328 – 1336.

R Development Core Team, 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL (http://www.R-project.org/). R version 2.14.

Reimann, C., Filzmoser, P., Garrett, R. and Dutter, R., 2008. Statistical data analysis explained: applied environmental statistics with R. John Wiley and Sons, Ltd, 2008.

Seroji, A.R., 2011. Particulates in the atmosphere of Makkah and Mina Valley during the Ramadan and Hajj seasons of 2004 and 2005. In Brebbia, C.A., Longhurst, J.W.S., and Popov, V., (ed) 2011. Air Pollution XIX, Wessex Institute of Technology, UK.

WHO, 2003. Health Aspects of Air Pollution with Particulate Matter, Ozone and Nitrogen Dioxide. Report on a WHO Working Group Bonn, Germany 13–15 January 2003.

WHO, 2004. World Health Organization, Protection of the human environment, assessing the environmental burden of disease at national and local levels, Geneva.