

# Modelling the non-linear association of particulate matter (PM<sub>10</sub>) with meteorological parameters and other air pollutants—a case study in Makkah

Said Munir<sup>1</sup>

Received: 31 July 2014 / Accepted: 6 November 2015 / Published online: 16 December 2015  
© Saudi Society for Geosciences 2015

**Abstract** Particulate matter with aerodynamic diameter of 10  $\mu\text{m}$  or less (PM<sub>10</sub>) causes numerous adverse health and environmental impacts; therefore, it is vital to characterise its behaviour in association with the controlling factors. In this paper, the effects of several meteorological parameters and gaseous pollutants on PM<sub>10</sub> concentrations ( $\mu\text{g}/\text{m}^3$ ) are analysed employing a quantile regression model (QRM). The study uses air quality and meteorology data collected in the arid climatic conditions of Makkah, Saudi Arabia. In this study, it is shown that the effects of covariates vary at different levels of PM<sub>10</sub> distributions, which confirms a non-linear association between PM<sub>10</sub> and independent variables. Temperature had positive significant effect at the middle quantiles (0.2 to 0.8) of PM<sub>10</sub>. The effect of atmospheric pressure was significant only at quantile 0.95 (slope =  $-1.85$ ). Relative humidity had significant effect at quantiles 0.05 to 0.3 and insignificant effect at higher quantiles. Both wind speed and lag\_PM<sub>10</sub> demonstrated significant positive effect at all quantiles, and the magnitude of slopes gradually increased as PM<sub>10</sub> concentration increased. The effect of CO was significant at all quantiles, and the magnitude of slopes ranged from  $-8$  to  $-47$  at quantile 0.05 and 0.95, respectively. The negative effect of SO<sub>2</sub> was significant at most of the quantiles, except at quantiles 0.05, 0.8 and 0.9, where the effect was insignificant. NO showed significant positive effect at all quantiles; in contrast, NO<sub>2</sub> had positive effect only at quantiles 0.05 to 0.6. The performance of the model was assessed both

locally (at each quantile) and globally (amalgamating the effect of all quantiles). QRM provides a new insight into air quality data analysis and outperforms other regression models.

**Keywords** PM<sub>10</sub> · Air pollution · Makkah—Saudi Arabia · Quantile regression model

## Introduction

It is well established that air pollution has negative impacts on human health, agricultural crops, ecosystem and building materials (e.g. Dockery et al. 1993; Burnett et al. 2000; WHO 2004; Al-Hobaib et al. 2010). Atmospheric particulate matters aggravate chronic respiratory and cardiovascular diseases, alter host defence, damage lung tissues, lead to premature death and possibly cause cancer (WHO 2004; Harrison 2001). 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 (AQEG 2009). Recently, several investigations have been made in Saudi Arabia and the surrounding Arab Gulf region to investigate spatial and temporal variability of particulate matters (Mashal et al. 2014; Munir et al. 2013b), quantify their emission sources (e.g. Abdelwaheb et al. 2014; Rushdi et al. 2013; Khodeir et al. 2012) and determine various factors contributing to the degradation of air quality (e.g., Al-Dabbas et al. 2013; Al-Khadouri et al. 2014). However, further investigations are required on health impacts and advanced modelling to

---

✉ Said Munir  
saidmunir@yahoo.co.uk

<sup>1</sup> Department of Environment and Health Research, The Custodian of the Two Holy Mosques Institute for Hajj and Umrah Research, Umm Al Qura University, Makkah, the Kingdom of Saudi Arabia

characterise the behaviour of aerosols, which can lead to better air quality management.

Makkah is one of the busiest cities in the world. Every year, millions of people visit the city due to its religious importance in the Muslim world. High levels of air pollutants is one of the growing concerns in Makkah, especially during the season of Hajj (Al-Jeelani 2009; Othman et al. 2010; Seroji 2011; Munir et al. 2013a; Munir et al. 2013b; Habeebullah 2013a; Habeebullah 2013b).  $PM_{10}$  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 by various atmospheric parameters, such as wind speed and direction, relative humidity, temperature and rainfall (e.g. Elminir 2005; Ordonez et al. 2005; Cheng et al. 2007; Beaver and Palazoglu 2009; Pearce et al. 2011). The levels of  $PM_{10}$  are associated with emission sources, the entering of particles from the ground surface, their residence time in the atmosphere, the formation of secondary pollutants, wind speed, turbulence level, air temperature and precipitation (Bhaskar and Mehta 2011). Furthermore, other air pollutants, for instance carbon monoxide (CO), sulphur dioxide ( $SO_2$ ) and nitrogen oxide ( $NO_x$ ), can affect  $PM_{10}$  concentrations in various ways. These pollutants can result in secondary aerosol 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 ( $(NH_4)_2SO_4$ ). Nitrogen dioxide ( $NO_2$ ) is oxidised to nitric acid ( $HNO_3$ ), which in turn can react with  $NH_3$  to form ammonium nitrate ( $NH_4NO_3$ ). The particles produced by the intermediate reactions of gases in the atmosphere are called secondary particles. 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. Furthermore, how the effects of meteorological parameters and other pollutants change at various regimes of the  $PM_{10}$  distributions require further considerations. Therefore, advanced modelling studies are required to analyse the effects of various controlling factors on  $PM_{10}$  to help better understand the association of these parameters with  $PM_{10}$ . This

paper intends to answer these questions, which is vital for preparing an effective management plan in Makkah and elsewhere.

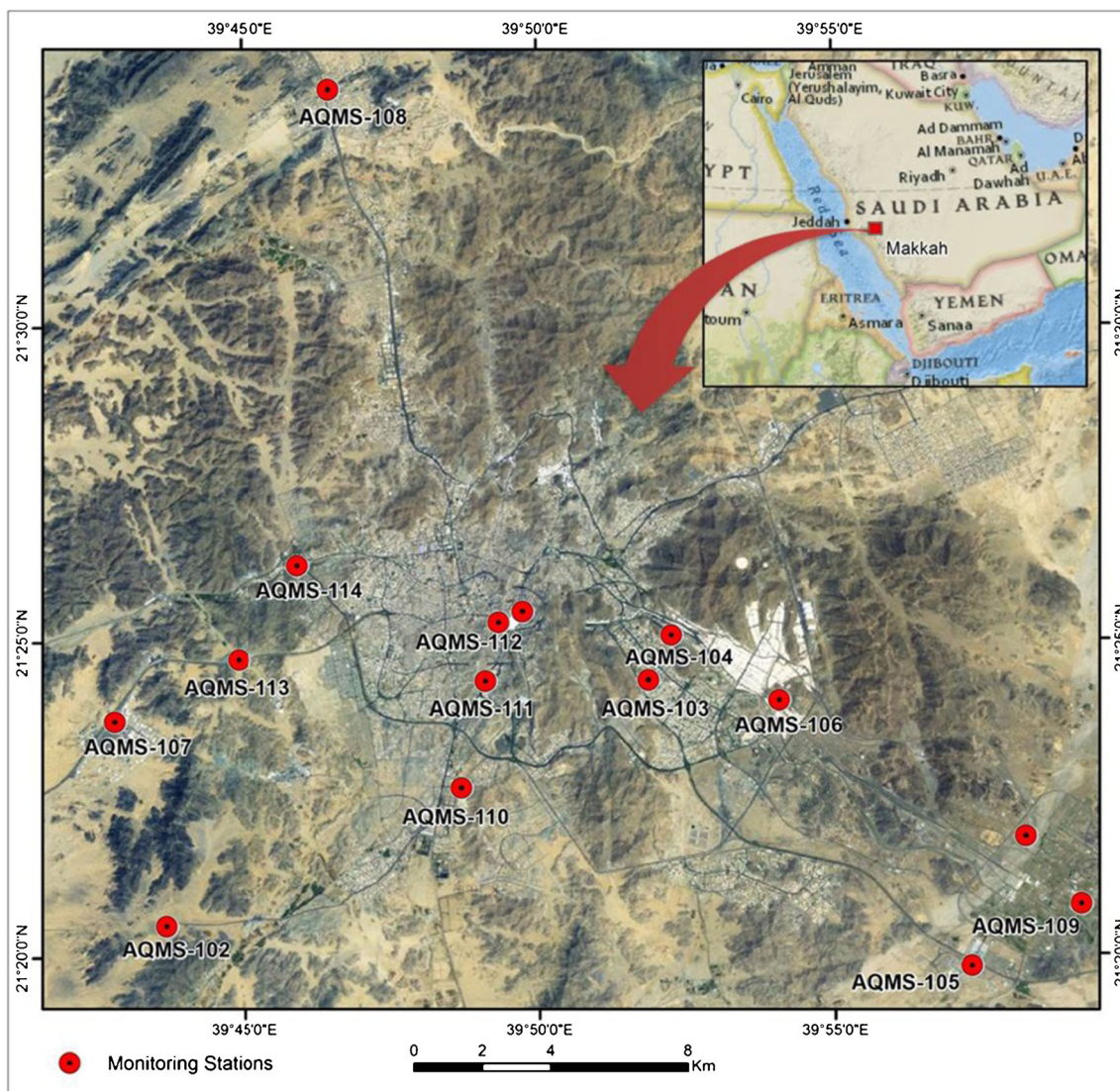
Previously Munir et al. (2013a) has developed a generalised additive model (GAM) to analyse the effect of several traffic-related air pollutants and meteorological parameters on  $PM_{10}$ ; however, GAM was unable to capture the variability in  $PM_{10}$  concentrations. Several metrics for the GAM model were estimated including coefficient of determination ( $R^2 = 0.52$ ), root mean square error (RMSE = 84) and fractional bias (FB = -0.22). In a more recent study by Sayegh et al. (2014), the performances of several statistical models were compared. QRM outperformed the other models and therefore was recommended for Makkah. However, there are several shortcomings in the Sayegh et al. (2014) study, which are addressed in this study: (a) data for the month of June were used as testing data, which for model comparison purposes is understandable as it would have the same effect on all models; however, ideally, the testing dataset should have been randomly selected. In this study the testing dataset has been randomly selected; (b) the outputs of QRM and its coefficients at different quantiles are not explained (for the purpose of brevity), which makes the study difficult to be understood, particularly for those readers who are new to this approach. In this study, the model outputs and how the coefficients change at various quantiles are explained in details; (c) only global prediction and resulted statistical metrics are considered, and no consideration is given to local performance of the model. This study analyses both global and local performance of the model. Furthermore, Sayegh et al. (2014) have focused on comparing the performance of several models, whereas this study focuses on the non-linear association between  $PM_{10}$  and the independent variables.

## Methodology

### Data source

This study uses data collected 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. This is a continuous monitoring station and measures several air pollutants and meteorological parameters. The location of the monitoring station is shown in Fig. 1. The air quality monitoring network was previously described by Munir et al. (2013a and b).

This study characterises  $PM_{10}$  concentration ( $\mu g/m^3$ ) with the aid of several air pollutants ( $CO$   $mg/m^3$ ,  $SO_2$   $\mu g/m^3$ ,  $NO$   $\mu g/m^3$ ,  $NO_2$   $\mu g/m^3$ ) and meteorological parameters (relative humidity (RH %), temperature ( $T$  °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



**Fig. 1** Map of the air quality and meteorological monitoring sites in Makkah, Saudi Arabia, where AQMS 112 represents the PME

these parameters is presented in Table 1, showing minimum (min), first quartile (0.25 quantile), mean, median (0.5 quantile), third quartile (0.75 quantile) and maximum levels of the given parameters. Data capture is greater than 90 % for all parameters, except SO<sub>2</sub> where 88 % data were present. 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<sup>3</sup> or mg/m<sup>3</sup>); however, PM<sub>10</sub> is always expressed as concentration (e.g. μg/m<sup>3</sup>). In this paper, all pollutants are expressed as concentrations (μg/m<sup>3</sup> or mg/m<sup>3</sup>) to be consistent in the use of units for both gaseous and non-gaseous pollutants.

It is shown in Fig. 2 that PM<sub>10</sub> 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) 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.

**General statistics**

Statistical Software R programming language (R Development Core Team 2012) and associated packages Quantreg, version 4.9.1 (Koenker 2012) and openair, version 2.13.2 (Carslaw and Ropkins 2012) were used for running



**Table 1** 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	First quartile	Mean	Median	Third quartile	Maximum	% Data capture
CO (mg/m <sup>3</sup> )	0	0.79	0.98	1.12	1.27	6.87	95
SO <sub>2</sub> (μg/m <sup>3</sup> )	0	5	8	11	15	125	88
NO <sub>2</sub> (μg/m <sup>3</sup> )	0	27	42	46	61	223	99
NO <sub>x</sub> (μg/m <sup>3</sup> )	0	21	33	42	52	367	99
NO (μg/m <sup>3</sup> )	0	2	5	12	13	299	99
PM <sub>10</sub> (μg/m <sup>3</sup> )	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 (°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

SO<sub>2</sub> sulphur dioxide, CO carbon monoxide, NO nitric oxide, NO<sub>2</sub> nitrogen dioxide, NO<sub>x</sub> nitrogen oxides, O<sub>3</sub> ozone, PM<sub>10</sub> particles with aerodynamic diameter of 10 μm or less, WS wind speed, WD wind direction, T temperature, RH relative humidity, P atmospheric pressure

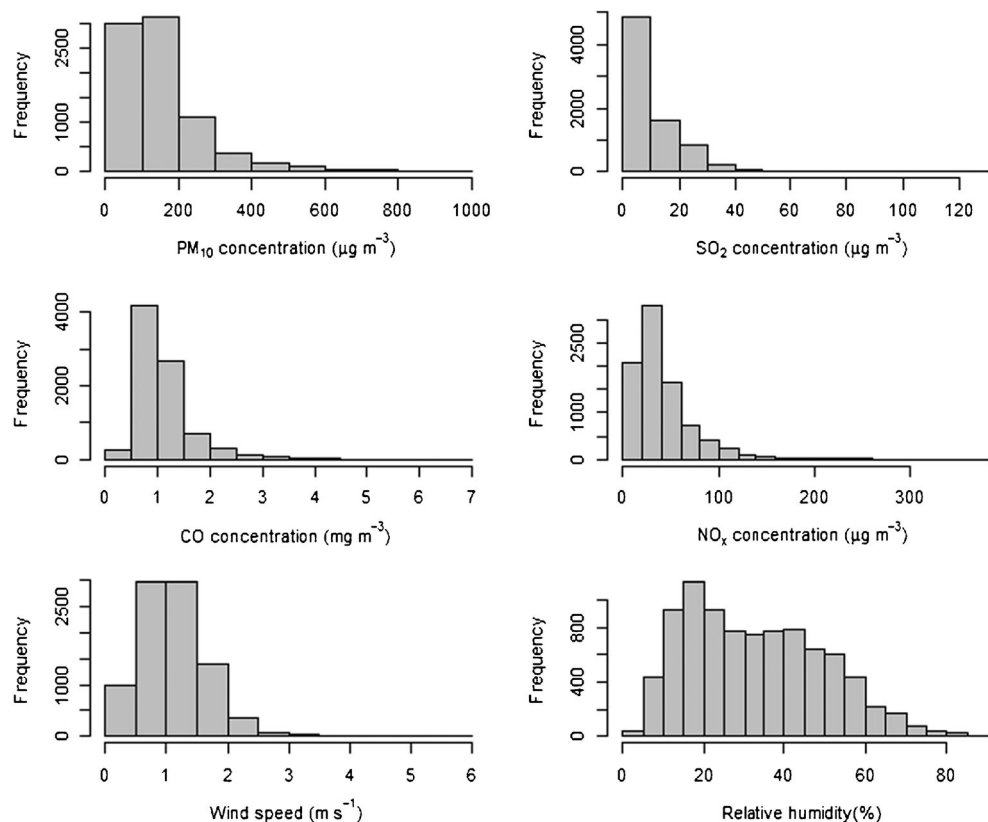
QRM, performing other statistical analysis and making graphs. Graphical presentations (e.g. histograms, polar plot and scatter diagram) are also used to present the outputs of the analysis.

### Quantile regression model

In this paper, QRM model is employed to analyse the effect of covariates (e.g. meteorological parameters and other air

pollutants such as NO<sub>x</sub>, CO, SO<sub>2</sub>) on PM<sub>10</sub> concentrations. QRM allows the covariates to have different contribution at different quantiles of the dependent variable distribution (here PM<sub>10</sub>) 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 Fig. 2. Furthermore, some air pollutants, such as ozone, exhibit non-linear association with its predictors (e.g. Gardner and Dorling

**Fig. 2** Histograms showing the frequency distributions of mean hourly data of PM<sub>10</sub>, SO<sub>2</sub>, CO, NO<sub>x</sub>, wind speed, and relative humidity at the PME monitoring station near the Holy Mosque in Makkah, Saudi Arabia for the year 2012



2000; Baur et al. 2004). This means that the contributions of the explanatory variables (e.g. meteorological variables) to independent variable vary significantly at different levels of ozone. This suggests that statistical models should have the capability to address the linearity and normality issues when applying to analyse air quality 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 concentrations in Athens, whereas Munir et al. (2012) modelled the effect of road traffic on ozone concentrations in the UK.

Using hourly mean PM<sub>10</sub> concentrations as a dependent (modelled or response) variable and several meteorological parameters (T, RH, P, WS and WD) and air pollutants (CO, NO, NO<sub>2</sub>, SO<sub>2</sub> and lag\_PM<sub>10</sub>) as independent variables, a QRM is developed intending to analyse the non-linear relationship between PM<sub>10</sub> the covariates. These covariates are important for modelling PM<sub>10</sub> concentrations and controlling a significant proportion of PM<sub>10</sub> variations as previously shown by Munir et al. (2013a). Multiple linear regression model (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). MLRM is used here to facilitate the understanding of QRM.

$$\text{PM}_{10} = \beta_0 + \beta_1 P + \beta_2 \text{RH} + \beta_3 T + \beta_4 \text{WS} + \beta_5 \text{WD} + \beta_6 \text{CO} + \beta_7 \text{SO}_2 + \beta_8 \text{NO} + \beta_9 \text{NO}_2 + \beta_{10} \text{lag\_PM}_{10} + \varepsilon_i \quad (1)$$

$$\text{PM}_{10} = \beta_0^{(p)} + \beta_1^{(p)} P + \beta_2^{(p)} \text{RH} + \beta_3^{(p)} T + \beta_4^{(p)} \text{WS} + \beta_5^{(p)} \text{WD} + \beta_6^{(p)} \text{CO} + \beta_7^{(p)} \text{SO}_2 + \beta_8^{(p)} \text{NO} + \beta_9^{(p)} \text{NO}_2 + \beta_{10}^{(p)} \text{lag\_PM}_{10} + \varepsilon_i \quad (2)$$

In equations (1) and (2),  $\beta_0$  represents the intercept,  $\beta_1$  to  $\beta_{10}$  represent the slopes (gradients) of the covariates and  $\varepsilon_i$  is 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 as follows: RMSE, normalised mean gross error (NMGE), coefficient of determination ( $R^2$ ), 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).

When we assess the performance of a model, we compare the observed concentrations with the predicted concentrations of the modelled variable, here PM<sub>10</sub>. Other statistical models, such as GAM or MLRM, have one prediction based on mean effect. In contrast, QRM has several predictions based on the number of quantiles used in the model. This makes assessing the performance of the QRM model somewhat different from other models. To assess the performance of QRM model, (1) either the prediction of QRM for each quantile is compared

with observed concentration or (2) global prediction (amalgamation of the prediction of all quantile) is compared with observed concentrations. The first method in which the prediction of each quantile is used is known as local performance, whereas the latter method is known as global performance.

To assess local performance, the prediction of each quantile was compared with observed PM<sub>10</sub> concentration of the test dataset. The test dataset was taken as 10 % independent random sample out of the total dataset. To evaluate the global performance of the QRM, this study adopts the amalgamated quantile regression model (AQRM) approach suggested by Baur et al. (2004). However, Baur et al. (2004) have used only  $R^2$  for assessing the model performance, whereas this paper extends this concept further to other metrics, such as NMB, NMGE, RMSE and FAC2. To determine these metrics, 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 predictions 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<sub>10</sub> concentration. The predicted PM<sub>10</sub> 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). ‘ModStat’ function in the openair package (Carslaw and Ropkins 2012) was used to calculate both local and global metrics for the model. An air quality model is considered acceptable if more than half of the predicted values are within a FAC2 of the observed concentrations 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).

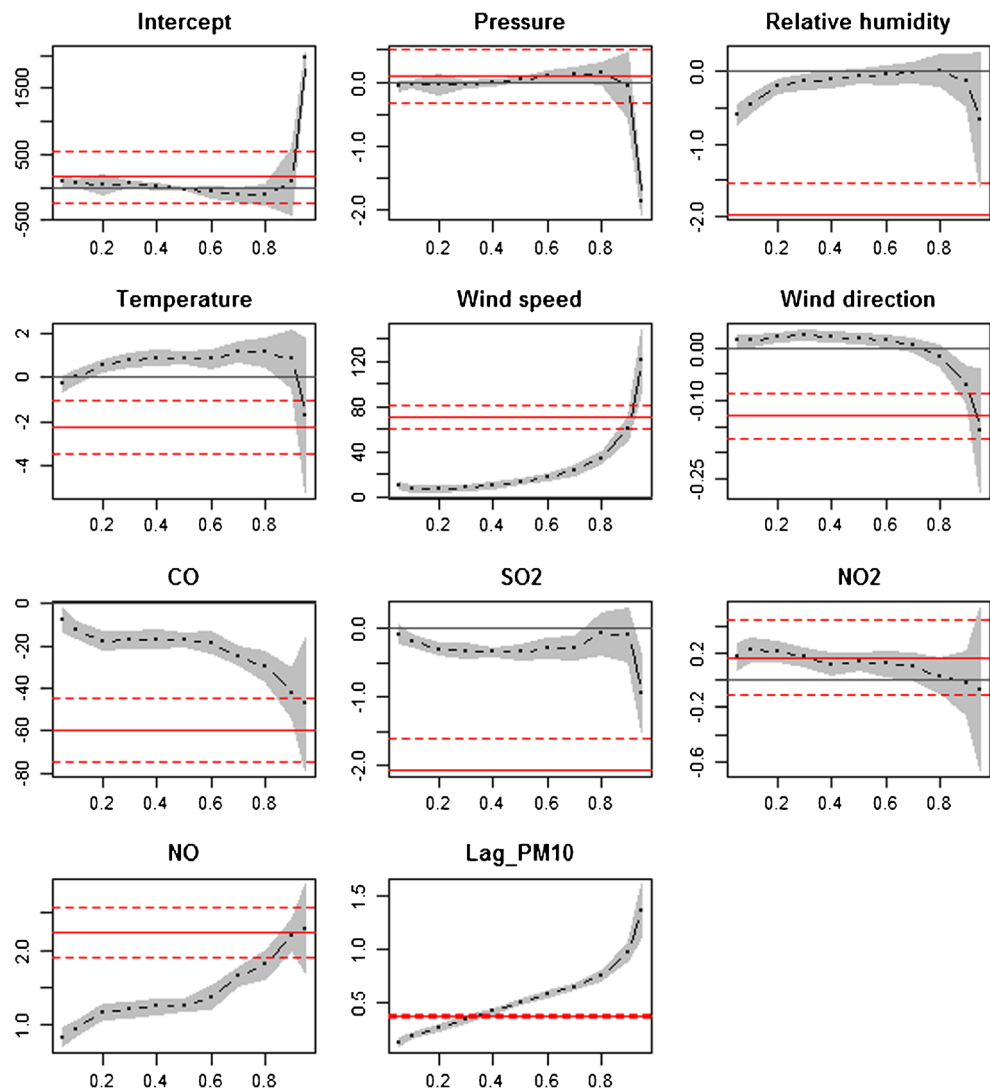
## Results and discussions

The outputs of QRM are depicted in Fig. 3, which shows the effect of various covariates on  $PM_{10}$  concentrations. The

quantiles used in this study are shown on the x-axis and their respective coefficients (slopes) are shown on the y-axis. The dashed-dotted black line represents the coefficients of QRM, the solid red line represents the mean coefficient and the solid black is the zero line. When confidence intervals overlap with the 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 Fig. 3 shows the intercept of the model. The intercepts are within the range of  $+100$  and  $-113$  for quantiles 0.9 and 0.8, respectively, except quantile 0.95 which has higher intercept. The effect of atmospheric pressure (Fig. 3, top-middle panel) is significant only at quantile 0.95, and for the rest of the quantiles, the confidence intervals overlap with the 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.

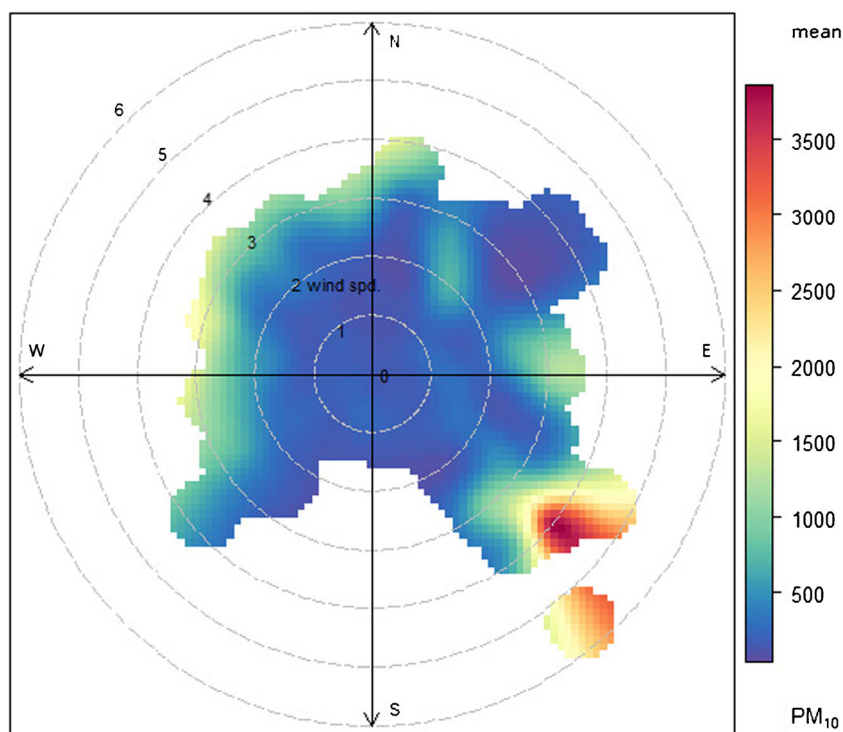
**Fig. 3** The outputs of quantile regression model (QRM) showing the effect of atmospheric pressure (hPa), relative humidity (%), temperature ( $^{\circ}C$ ), wind speed (m/s), wind direction (degrees from the north), carbon monoxide ( $CO$   $mg/m^3$ ), sulphur dioxide ( $SO_2$   $mg/m^3$ ), nitrogen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



This means that 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. 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 non-significant at higher quantiles. As reported previously by Munir et al. (2013a), high relative humidity is generally linked with night times 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) have 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  $NO_2$  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 result in enhanced re-suspension of soil and road dust and formation of secondary aerosol; hence, a temperature increase from 10 to 35 °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 are mostly caused by sand storms and construction activities near the monitoring site (Munir et al. 2013b), which are more dependent on wind speed and direction than temperature; therefore, probably that is why temperature shows non-significant effect. The mean effect of temperature is negative, and the regression coefficient is about  $-2$ . Mean can be biased by the presence of outliers in the data. Therefore, for air quality analysis, more robust metrics (e.g. median or other quantiles) should be used, which are not affected by extreme values. When temperature was used as the only model input, even the mean 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. Wind speed shows much stronger effect than the other covariates. The effect gradually becomes stronger as  $PM_{10}$  concentration increases, showing greater rate of increase at higher quantiles. The slope for wind speed at quantile 0.95 is about 120. 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 deserts around the Makkah city

**Fig. 4** Polar plot of  $PM_{10}$  concentration ( $\mu g/m^3$ ) near the Holy Mosque, Makkah, colour-coded by  $PM_{10}$  concentrations for 2012

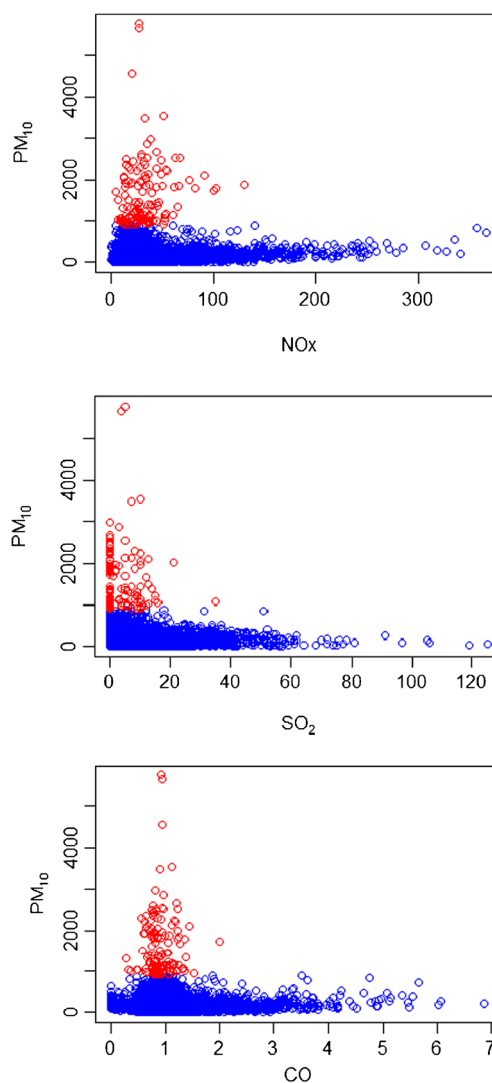


causing sand and dust storms. This might show that the sources of  $PM_{10}$  are mostly regional. In case of local sources, the wind speed would normally have negative impact by dispersing the locally emitted particles (e.g. Barmpadimos et al. 2011). The effect of wind direction is positive at lower 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 (Fig. 4).

Polar plots are constructed by averaging pollutant concentrations 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 GAM smoothing techniques (Carslaw and Ropkins 2012). It can be observed in Fig. 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, high  $PM_{10}$  concentration is shown in the west, northwest and east directions. Mostly low  $PM_{10}$  concentration can be observed at low wind speed (<2 m/s) from all directions. 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 the 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  $PM_{10}$  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 slopes range from  $-8$  to  $-47$  at quantiles 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 quantiles 0.9 and 0.95. The effect of  $SO_2$  is negative and significant at most of the quantiles, except at quantiles 0.05, 0.8 and 0.9. Mean regression coefficient is about  $-2$  and is significantly different from the quantile regression coefficients. The positive effect of  $NO_2$  is significant at quantiles 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, for NO, the strength of coefficients gradually increases from quantiles 0.05 to 0.95, in contrast to  $NO_2$ , where the strength of coefficients shows the opposite pattern. The effect of lag\_  $PM_{10}$  (previous day  $PM_{10}$  concentration) is positive, 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\_  $PM_{10}$  demonstrates positive effect.

It can be observed in Fig. 3 that the effect of independent variables on  $PM_{10}$  concentration is not linear and changes as the concentration of  $PM_{10}$  changes. For some variables, 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 for other covariates, 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 as in the case of relative humidity, temperature,  $NO_2$  and  $SO_2$ ; however, temperature,



**Fig. 5** Scatter plots of hourly  $PM_{10}$  concentrations ( $\mu\text{g}/\text{m}^3$ ) versus NOx ( $\mu\text{g}/\text{m}^3$ ), CO ( $\text{mg}/\text{m}^3$ ) and  $SO_2$  ( $\mu\text{g}/\text{m}^3$ ) concentrations measured at PME monitoring stations near the Holy Mosque in Makkah, Saudi Arabia, 2012. The red and blue colour indicates different patterns in the association of  $PM_{10}$  and the gaseous pollutants



pressure and SO<sub>2</sub> show insignificant effect at lower quantiles as well. This type of relationship usually remains hidden when applying linear models, e.g. MLRM, which assumes linear association between dependent and independent variables.

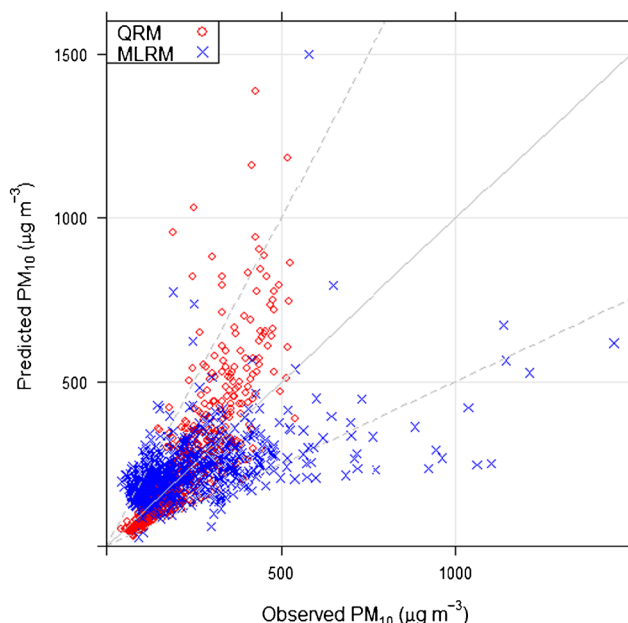
CO and SO<sub>2</sub> would be expected to show positive association with PM<sub>10</sub> concentrations if they had the same sources of emissions. However, here the association is predominantly negative, which probably shows that they have different sources of emissions. In Makkah, PM<sub>10</sub> mainly comes from re-suspension and windblown dust and sand particles, whereas the gaseous pollutants are mainly emitted by road traffic (e.g. Habeebullah 2013a; Munir et al. 2013 a and b). In addition, meteorological parameters, especially wind speed, probably play an important role in the negative association of PM<sub>10</sub> and gaseous air pollutants. To investigate this further, scatter plots of CO, SO<sub>2</sub> and NO<sub>x</sub> against PM<sub>10</sub> are shown in Fig. 5, which clearly shows two different patterns in the association of PM<sub>10</sub> and gaseous pollutants. The red colour shows high PM<sub>10</sub> concentrations associated with low concentrations of gaseous pollutants (e.g. CO). The blue colour indicates a different pattern; i.e., as the concentrations of gaseous pollutants increase, PM<sub>10</sub> concentrations show little variations. Wind speed probably plays the dominant role in the negative association of PM<sub>10</sub> with the gaseous air pollutants. High wind speed, on the one hand, blowing sand and dust particles, enhances the concentration of PM<sub>10</sub>; on the other hand, dispersing locally emitted gaseous pollutants, it reduces the concentrations of gaseous pollutants. PM<sub>10</sub> levels are extremely high in the red sections, showing extreme episodes of PM<sub>10</sub>, which are probably caused by wind storms in Makkah. Episodes of high PM<sub>10</sub> are associated with low levels of other pollutants and vice versa, which probably explains the negative effect of CO and SO<sub>2</sub> on PM<sub>10</sub> concentration.

**Table 2** Statistical metrics for assessing the global performance of the model calculated for the testing dataset (10 % random sample) at PME monitoring station in Makkah

Metric	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 <sup>2</sup>	0.82	0.39

For definitions and calculation methods of these metrics, see Derwent et al. (2010) and Carslaw (2011)

QRM quantile regression model, MLRM multiple regression model, FAC2 factor of 2, MB mean bias, RMSE root mean square error, MGE mean gross error, NMGE normalised mean gross error, R<sup>2</sup> for coefficient of determination, NMB normalised mean bias



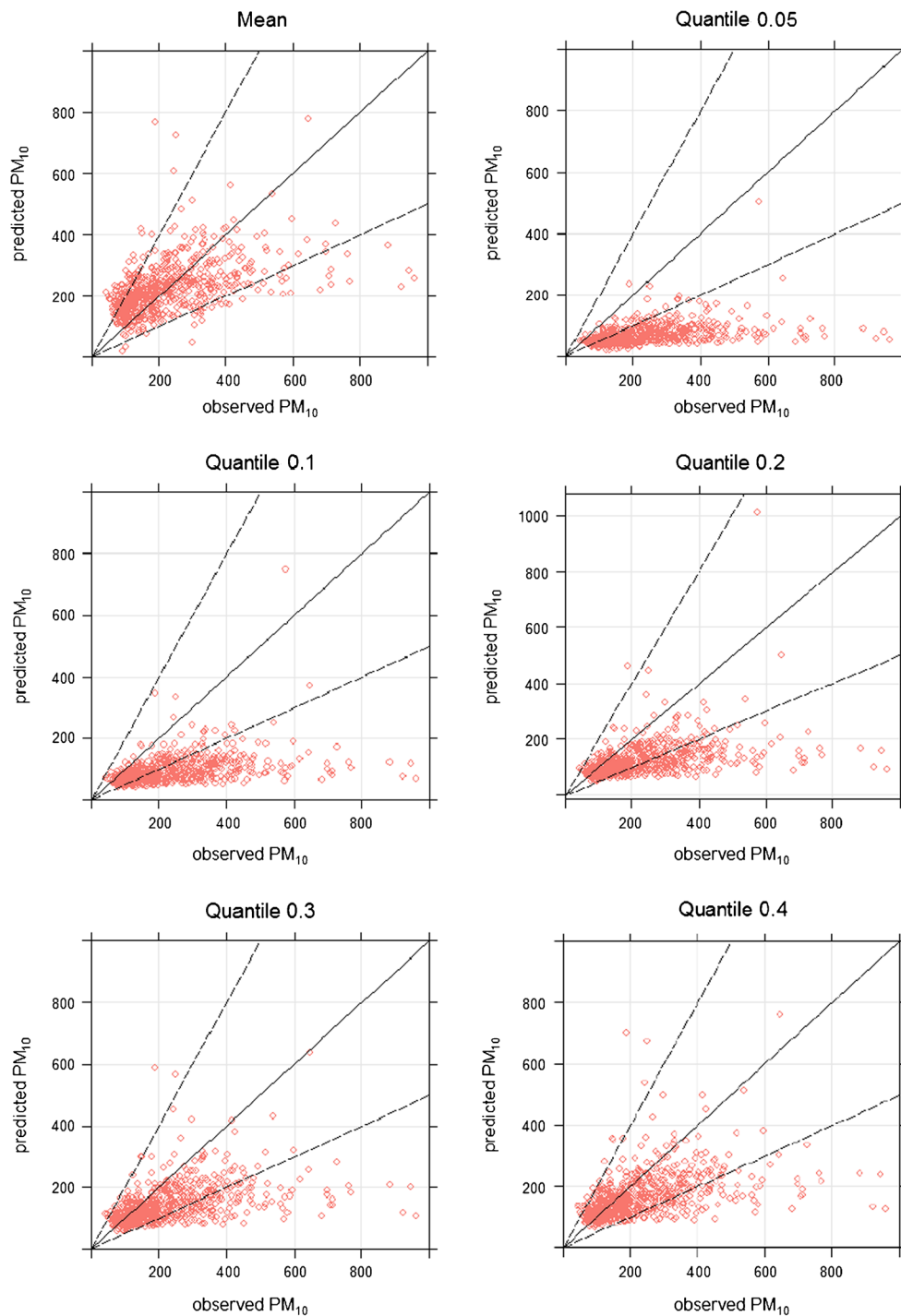
**Fig. 6** Comparison of observed and predicted PM<sub>10</sub> concentrations (µg/m<sup>3</sup>) based on the testing dataset for 2012. The middle solid line is 1:1, and the above and below dashed lines are 0.5:1 and 2:1, respectively. So, the area between the two dashed lines is the factor of two (FAC2) regions

### QRM model assessment

The performance of QRM was assessed by both using global prediction and local prediction of each quantile. Firstly, the data were divided into two subsets: training data and testing data. For testing data, a 10 % random sample was selected, which was not included in the training dataset. Table 2 shows the values of various metrics calculated from global prediction (as described in “Quantile regression model” section). It is shown in Table 2 that the values of these metrics for both QRM and MLRM are within the recommended range, as

**Table 3** Statistical metrics for assessing the local performance at each quantile of PM<sub>10</sub> concentrations (µg/m<sup>3</sup>). The metrics are calculated for the testing dataset (10 % random sample) at PME monitoring station in Makkah

Quantile	FAC2	MB	MGE	NMB	NMGE	RMSE	R <sup>2</sup>
0.05	0.18	-175.06	175.26	-0.72	0.72	274.92	0.34
0.1	0.41	-150.25	152.21	-0.62	0.63	257.37	0.36
0.2	0.64	-121.53	128.47	-0.50	0.53	239.44	0.37
0.3	0.75	-96.94	112.33	-0.40	0.46	226.62	0.38
0.4	0.82	-74.52	102.21	-0.31	0.42	217.84	0.38
0.5	0.85	-51.95	98.45	-0.21	0.41	212.97	0.38
0.6	0.87	-28.75	98.18	-0.12	0.41	211.37	0.38
0.7	0.85	-0.05	102.35	-0.00	0.42	212.89	0.39
0.8	0.81	40.04	116.19	0.16	0.48	223.87	0.39
0.9	0.63	118.18	165.04	0.49	0.68	270.39	0.40
0.95	0.39	254.18	278.84	1.05	1.15	396.28	0.40



**Fig. 7** Scatter plots of predicted and observed PM<sub>10</sub> concentrations ( $\mu\text{g}/\text{m}^3$ ) at various quantiles based on the testing dataset (10 % random sample) for 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

more than half of the predicted values are within a FAC2 of the observed concentration and NMB values lie within the range of  $-0.2$  and  $+0.2$  (Derwent et al. 2010). Therefore, the performance of the models is acceptable. In addition, the performance of the QRM is better than that of MLRM; for instance,

FAC2 and  $R^2$  for QRM and MLRM are 0.96, 0.82 and 0.82, 0.39, respectively.

Figure 6 compares observed and predicted PM<sub>10</sub> concentrations of both QRM and MLRM with the help of a scatter plot, which is very useful for model evaluation (Carslaw

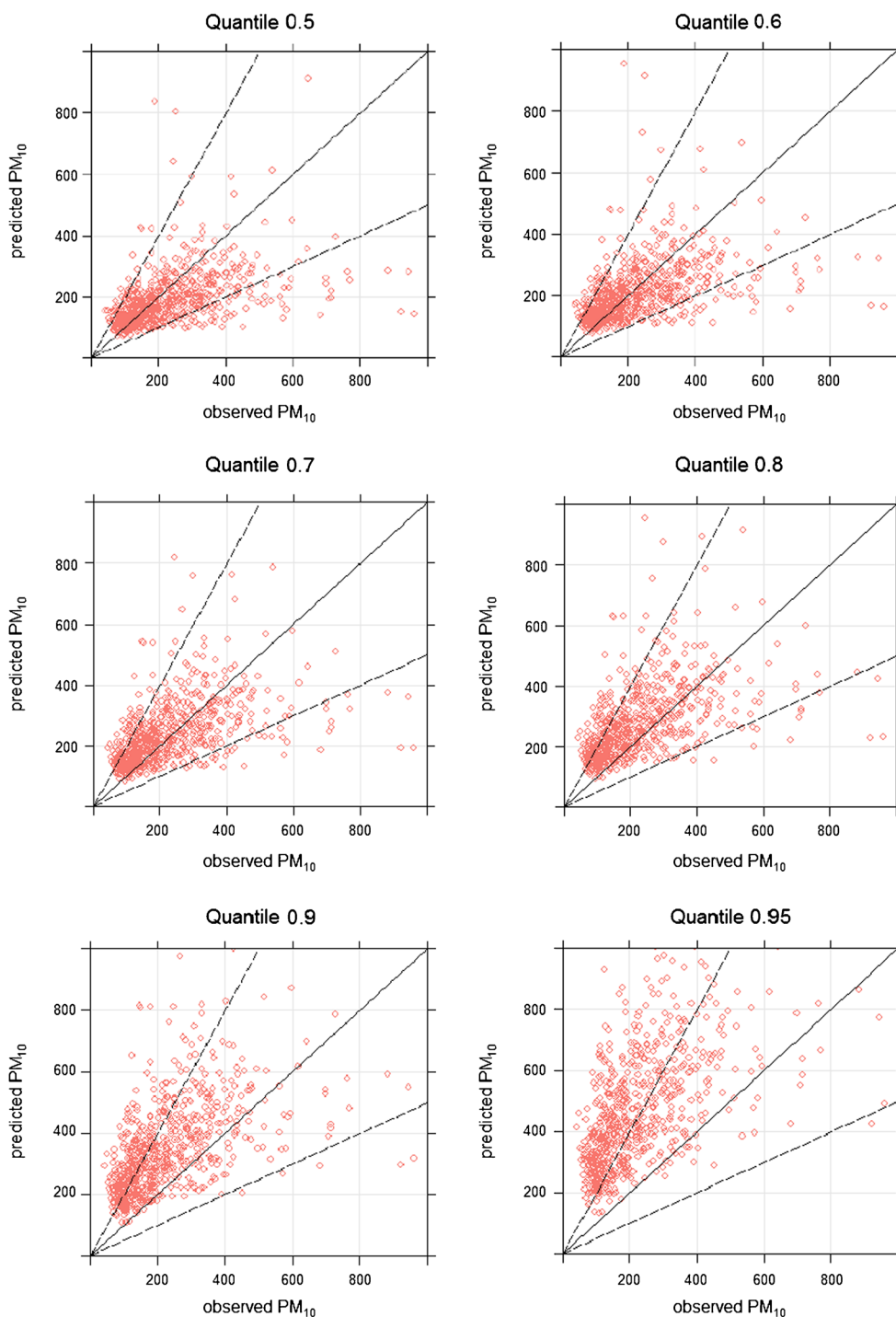


Fig. 7 (continued)

2011). In the scatter plot, it is much easier to see where the data lie and to get a feeling about bias. Relatively, more points lie below the 1:1 line (middle line in Fig. 6) in the case of MLRM and there seems to be a slight negative bias (under prediction), whereas more points lie above the 1:1 line in the case of QRM, showing slight positive bias (over prediction). Particularly, at high concentration of PM<sub>10</sub>, MLRM fails to perform and under-predicts PM<sub>10</sub> 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 3 shows various metrics calculated for each quantile to show local performance of the model at each quantile using testing dataset (10 % independent random sample). The values of FAC2 show that the model performance is acceptable at all quantiles, except at both tails of the distribution, i.e. quantiles

0.05, 0.1 and 0.95. At these quantiles, less than half of the predicted values are within a FAC2 of the observed concentrations ( $FAC2 < 0.50$ ). The greatest FAC2 value is shown by quantile 0.6 ( $FAC2 = 0.87$ ), followed by quantiles 0.5 and 0.7 both having FAC2 value of 0.85. The values of NMB are between +0.2 and -0.2 at quantiles 0.6, 0.7 and 0.8. Most of the metrics show best performance either at quantile 0.6 or 0.7, except  $R^2$  which shows the highest value at quantiles 0.9 and 0.95 ( $R^2 = 0.40$ ). The scatter plots (Fig. 7) compare the observed and predicted  $PM_{10}$  concentrations at various quantiles of  $PM_{10}$ . It can be clearly observed in the scatter plots that the model under-predicts  $PM_{10}$  concentrations at lower quantiles (quantiles 0.05 to 0.4) where most of the points lie below the 1:1 line. On the other hand, the model over-predicts  $PM_{10}$  concentrations at the higher quantiles (0.8, 0.9 and 0.95), and most of the points lie above the 1:1 line. When the predictions of all quantiles are integrated away as described in “Quantile regression model” section, the performance of the model significantly improves as shown in Fig. 6 and Table 2.

## Conclusions

This study employs a QRM to characterise the effect of several air pollutants and meteorological variables on  $PM_{10}$  concentrations in Makkah, Saudi Arabia. QRM characterises the effect of covariates at various quantiles, in contrast to the traditional approaches which analyse 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<sub>2</sub>, NO, NO<sub>2</sub> and lag\_  $PM_{10}$ ) was significant in at least 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. It is shown that the effect is non-linear and changes with the levels of  $PM_{10}$  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 for both global and local predictions. Global prediction shows much better performance than prediction for each individual quantile. The middle quantiles (0.5, 0.6 and 0.7) showed better performance than tails at both ends. Further investigations are required to identify various sources of  $PM_{10}$  and quantify their contributions to the observed  $PM_{10}$  concentrations, including road traffic in Makkah which is part of the ongoing project for improving air quality in Makkah.

**Acknowledgments** 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

- Abdelwaheb A, Moncef Z, Hamed BD (2014) Quantification of air pollutant emissions from the Jebel Chakir landfill site (Tunis city). Arab J Geosci. doi:10.1007/s12517-014-1350-x
- Al-Dabbas MA, Ali LA, Afaj AH (2013) Determination of total suspended particles and the polycyclic aromatic hydrocarbons concentrations in air of selected locations at Kirkuk. Iraq Arab J Geosci doi. doi:10.1007/s12517-013-1229-2
- Al-Hobaib AS, Al-Jaseem KQ, Baioumy HM, Ahmed AH (2010) Environmental impact assessment inside and around Mahd Adh Dahab gold mine, Saudi Arabia. Arab J Geosci 5:985–997. doi:10.1007/s12517-010-0259-2
- Al-Jeelani HA (2009) Evaluation of air quality in the holy Makkah during Hajj season 1425 H. J Appl Sci Res 5:115–121
- Al-Khadouri A, Al-Yahya S, Charab Y (2014) Contribution of atmospheric processes to the degradation of air quality: case study (sohar industrial area, Oman). Arab J Geosci. doi:10.1007/s12517-014-1295-0
- 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
- Barnpadimos I, Hueglin C, Keller J, Henne S, Prevot A (2011) Influence of meteorology on  $PM_{10}$  trends and variability in Switzerland from 1991 to 2008. Atmos Chem Phys 11:1813–1835
- Baur D, Saisana M, Schulze N (2004) Modelling the effects of meteorological variables on ozone concentration—a quantile regression approach. Atmos Environ 38:4689–4699
- Beaver S, Palazoglu A (2009) Influence of synoptic and mesoscale meteorology on ozone pollution potential for San Joaquin Valley of California. Atmos Environ 43(10):1779–1788
- Bhaskar BV, Mehta VM (2011) Atmospheric particulate pollutants and their relationship with meteorology in Ahmed Abad. Aerosol Air Qual Res 10:301–315
- Burnett RT, Brook J, Dann T, Delocla C, Philips O, Cakmak S, Vincent R, Goldberg MS, Krewski D (2000) Association between particulate- and gas-phase components of urban air pollution and daily mortality in eight Canadian cities. Inhal Toxicol 12(Suppl. 4):15–39
- Carlsaw D (2011) Defra regional and transboundary model evaluation analysis—phase 1. Version: 15th April 2011
- Carlsaw D, Ropkins K (2012) Openair—an R package for air quality data analysis. Environ Model Softw 27:52–61
- Cheng CSQ, Campbell M, Li Q, Li G, Auld H, Day N, Pengelly D, Gingrich S, Yap D (2007) A synoptic climatological approach to assess climatic impact on air quality in south-central Canada—part I: historical analysis. Water Air Soil Poll 182(1–4):131–148
- Derwent D, Fraser A, Abbott J, Jenkin M, Willis P, Murrells T (2010) Evaluating the performance of air quality models. Issue 3/June 2010.7, 81
- Dockery DW, Pope CA, Xu X, Spengler JD, Ware JH, Fay ME, Ferris BG, Speizer FE (1993) An association between air pollution and mortality in six US cities. N Engl J Med 329:1753–1759
- Duenas C, Fernandez MC, Canete S, Carretero J, Liger E (2002) Assessment of ozone variations and meteorological effects in an urban area in the Mediterranean coast. Sci Total Environ 299:97–113
- Elminir HK (2005) Dependence of urban air pollutants on meteorology. Sci Total Environ 350(1–3):225–237



- Gardner MW, Dorling SR (2000) Statistical surface ozone models: an improved methodology to account for non-linear behaviour. *Atmos Environ* 34:21–34
- Habeebullah TM (2013a) Health impacts of PM<sub>10</sub> using AirQ2.2.3 model in Makkah. *J Basic Appl Sci* 9:259–268
- Habeebullah TM (2013b) Risk assessment of exposure to BTEX in the Holy City of Makkah. *Arab J Geosci* doi 10.1007/s12517-013-1231-8
- Hao L, Naiman DQ (2007) *Quantile regression: series-quantitative applications in the social sciences*. Sage Publications, 2007 (Series NO. 07–149)
- Harrison RM (2001) *Pollution causes, effects, and control*. Royal Society of Chemistry, Fourth Edition
- Khodeir M, Shamy M, Alghamdi M, Zhong M, Sun H, Costa M, Chen LC, Maciejczyk PM (2012) Source apportionment and elemental composition of PM<sub>2.5</sub> and PM<sub>10</sub> in Jeddah city, Saudi Arabia. *Atmos Poll Res* 3:331–340
- Koenker R (2005) *Quantile regression, economic society monographs* No.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>)
- Mashal K, Salahat M, Al-Qinna M, Al-Degs Y (2014) Spatial distribution of cadmium concentrations in street dust in an arid environment. *Arab J Geosci*. doi:10.1007/s12517-014-1367-1
- 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 CA, Longhurst JWS, Popov V (Eds.), 2011. *Air Poll XIX*, 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. *Atmos Environ* 60:283–291
- Munir S, Habeebullah TM, Seroji AR, Morsy EA, Mohammed AMF, Saud WA, Esawee AL, Awad AH (2013a) Modelling particulate matter concentrations in Makkah, applying a statistical modelling approach. *Aerosols Air Qual Res* 13:901–910
- Munir S, Habeebullah TM, Seroji AR, Gabr SS, Mohammed AMF, Morsy EA (2013b) Quantifying temporal trends of atmospheric pollutants in Makkah (1997–2012). *Atmos Environ* 77:647–655
- Ordonez C, Mathis H, Furger M, Henne S, Huglin C, Staehelin J, Prevot ASH (2005) Changes of daily surface ozone maxima in Switzerland in all seasons from 1992 to 2002 and discussion of summer 2003. *Atmos Chem Phys* 5:1187–1203
- Othman N, Mat-Jafri MZ, San LH (2010) Estimating particulate matter concentration over arid region using satellite remote sensing: a case study in Makkah, Saudi Arabia. *Mod Appl Sci* 4:11–20
- Pearce JL, Beringer J, Nicholls N, Hyndman RJ, Tapper NJ (2011) Quantifying the influence of local meteorology on air quality using generalized additive models. *Atmos Environ* 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 (2008) *Dutter R (2008) statistical data analysis explained: applied environmental statistics with R*. John Wiley and Sons, Ltd
- Rushdi AI, Al-Mutlaq KF, Al-Otaibi M, El-Mubarak AH, Simoneit BRT (2013) Air quality and elemental enrichment factors of aerosol particulate matter in Riyadh City, Saudi Arabia. *Arab J Geosci* 6:585–599. doi:10.1007/s12517-011-0357-9
- Sayegh AS, Munir S, Habeebullah TM (2014) Comparing the performance of statistical models for predicting PM<sub>10</sub> concentrations. *Aerosol Air Qual Res* 14:653–665
- Seroji AR (2011) Particulates in the atmosphere of Makkah and Mina Valley during the Ramadan and Hajj seasons of 2004 and 2005. In Brebbia CA, Longhurst JWS and Popov V (ed) 2011. *Air poll XIX*, Wessex Institute of Technology, WIT press Southampton, Boston
- 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