

PAPER • OPEN ACCESS

## Predicting Particulate Matter (PM<sub>10</sub>) during High Particulate Event (HPE) using Quantile Regression in Klang Valley, Malaysia

To cite this article: N A A A Rahim *et al* 2023 *IOP Conf. Ser.: Earth Environ. Sci.* **1216** 012003

View the [article online](#) for updates and enhancements.

You may also like

- [An LSTM-based neural network method of particulate pollution forecast in China](#)  
Yarong Chen, Shuhang Cui, Panyi Chen et al.
- [Analysis of PM<sub>10</sub> in urban and rural environment in Sumatra Island over the past half-decade](#)  
Aulia N Khoir, Suradi, Kurdiyan et al.
- [What matters in public perception and awareness of air quality? Quantitative assessment using internet search volume data](#)  
Young-Hee Ryu and Seung-Ki Min



**ECS** The Electrochemical Society  
Advancing solid state & electrochemical science & technology

**250**  
ECS MEETING CELEBRATION

*Step into the  
Spotlight*

**SUBMIT YOUR  
ABSTRACT**

**250th ECS Meeting**  
**October 25–29, 2026**  
**Calgary, Canada**  
*BMO Center*

*Submission deadline:*  
**March 27, 2026**

# Predicting Particulate Matter (PM<sub>10</sub>) during High Particulate Event (HPE) using Quantile Regression in Klang Valley, Malaysia

N A A A Rahim<sup>1,2</sup>, N Mohamed Noor<sup>1,2,\*</sup>, I A Mohd Jafri<sup>1,2</sup>, N Ramli<sup>1,2</sup>, M A Kamaruddin<sup>3</sup>, G Deák<sup>4</sup>

<sup>1</sup>Faculty of Civil Engineering Technology, Universiti Malaysia Perlis, Kompleks Pengajian Jejawi 3, 02600 Arau, Perlis, Malaysia

<sup>2</sup>Sustainable Environment Research Group (SERG), Centre of Excellence Geopolymer and Green Technologies (CEGeoGTech), Universiti Malaysia Perlis, Kompleks Pengajian Jejawi 2, 02600 Arau, Perlis, Malaysia

<sup>3</sup>School of Industrial Technology, Universiti Sains Malaysia, 11800 Pulau Pinang, Malaysia

<sup>4</sup>National Institute for Research and Development in Environmental Protection Bucharest (INCDPM), 294 Splaiul Independentei Street, 6th District, 060031 Bucharest, Romania

\*norazian@unimap.edu.my

**Abstract.** Particulate matter (PM<sub>10</sub>) is the key indicator of air quality index (API) during high particulate event (HPE). The presence of PM<sub>10</sub> is believed to have an adverse effect on human health and environment. Therefore, the prediction of future PM<sub>10</sub> concentration is very important because it can aid the local authorities to implement precautionary actions to limit the impact of air pollution. This study aims to compare the performances of two predictive models, which include Multiple Linear Regression (MLR) and Quantile Regression (QR) in predicting the next-day PM<sub>10</sub> concentration during HPE. The hourly dataset of PM<sub>10</sub> concentration with other trace gases and weather parameters at Kelang and Petaling Jaya from the year of historic haze event in Malaysia (1997, 2005, 2013 and 2015) were obtained from Department of Environment (DOE) Malaysia. Three performance measures namely Mean Absolute Error (MAE), Normalised Absolute Error (NAE) and Root Mean Squared Error (RMSE) were calculated to evaluate the performances of the predictive models. From the results, QR model at quantile 0.3 and 0.6 was chosen as the best predictive tools for predicting the next day PM<sub>10</sub> concentration during haze event in Kelang and Petaling Jaya, respectively. showed better performance for the prediction of next-day PM<sub>10</sub> concentration in Kelang. These results indicate that QR can be used as one of predictive tool to forecast air pollution concentration especially during unusual condition of air quality.

## 1. Introduction

Air pollution has come to be a major problem in Malaysia for over a decade. One of the significant air pollution in Malaysia was haze that usually transported from the neighbouring country. Haze, which also known as high particulate event (HPE) refers to a meteorological condition in which the presence



of suspended solid or liquid particles, smoke, and vapour causes atmospheric visibility to be less than 10 kilometres [1]. The formation of HPE has been linked with the meteorological conditions as well as anthropogenic activities such as industrial activities, mobile emissions and biomass burning [2–4]. HPE has deteriorated the local air quality, which contributed to high Air Pollution Index (API) reading in Malaysia. The first HPE incident in Malaysia was recorded in the 1980s [5]. Particulate matter is the significant criterion of air quality index in Malaysia and Southeast Asia's main haze-related pollutant. Particulate matter with an aerodynamic diameter of less than 10 µm or known as PM<sub>10</sub> is a combination of carbon, dust and aerosols. Several HPE occurrences in Malaysia have been recorded where the PM<sub>10</sub> concentrations surpassed the Recommended Malaysian Ambient Air Quality Guideline (150 µg/m<sup>3</sup> for a 24-hour average) at one or more stations. The continuous HPE episodes has badly affected the human health as well as the overall environment. Therefore, it is very important to understand the relationship and assess and monitor the PM<sub>10</sub> levels by using a forecasting model in order to improve air quality.

Multiple statistical approaches have been applied in air pollution studies in Malaysia. Multiple Linear Regression (MLR) is one of the practices that has been commonly used in forecasting air pollution [6]. Several studies on the prediction model for PM<sub>10</sub> concentration, precisely in the East Coast of Peninsular Malaysia have been conducted by developing the MLR model, based on several site classifications and during different monsoon season to determine its variation during non-haze periods [7–9]. However, MLR has its own limitation due its failure to encompass the response to non-central location of predatory variables and its incapability to satisfy with the model assumptions [10]. Oppositely, a study by [11] found that non-linear and learning machine means outperformed the linear regression methods. The technique is simple and easy to be deployed thus it is still widely used.

Quantile regression (QR) is another method that has been used in predicting the PM<sub>10</sub> concentrations. It is progressively developing as a thorough method to the statistical analysis of linear and nonlinear models [11]. The QR represented the non-central location of a distribution which allow the approach to be more useful and precise [12]. [13] concluded that QR models inhibit some advantages compared to MLR since it does not rely on any properties, is independent or only mildly dependent, is robust to outliers, and is distribution free. A study by [14] found that the application of QR was able to deal with the influences of meteorological variables on ozone (O<sub>3</sub>) levels in Hong Kong at various percentiles. Research conducted by [11] found that the influence of independent variables varies depending on the O<sub>3</sub> quantile distributions and that QR was able to delineate the nonlinear relationship between O<sub>3</sub> and the independent variables in Athens. [10] found out that QR indicated better performance in predicting the future PM<sub>10</sub> concentrations in Seberang Perai, Malaysia, by comparing the performance of QR and MLR.

The goal of this paper is to compare the performance of MLR and QR models in predicting the next-day PM<sub>10</sub> levels in Kelang and Petaling Jaya. The finding of this study would be useful for the authority in planning and carrying out required actions to limit their exposure to air pollution and improving air quality on selected locations.

## 2. Materials and method

### 2.1. Study location

The selected study areas are located in the west coast region of Peninsular Malaysia. The specific locations and its descriptions were tabulated in Table 1.

**Table 1.** Specific location of air quality monitoring stations in Klang Valley.

Station Location	Coordinates		Background
	Latitude (E)	Longitude (N)	

Sekolah Menengah (P) Raja Zarina	3.01217°	101.40978°	Urban
Sekolah Kebangsaan Bandar Utama	3.13217°	101.60814°	Urban

2.2. Air Quality Dataset

The dataset were obtained from Department of Environment (DOE), Malaysia and the data includes of hourly data of particulate matter with less than 10 microns in size (PM<sub>10</sub>); traces gases including nitrogen oxides (NO<sub>x</sub>), sulphur dioxide (SO<sub>2</sub>), nitrogen dioxide (NO<sub>2</sub>), ozone (O<sub>3</sub>), carbon monoxide (CO); and weather parameters such as wind speed (WS), ambient temperature (T) and relative humidity (RH) during the year Malaysia experienced historic haze event (1997, 2005, 2013 and 2015) were used in this study. The parameters involved and its units were tabulated in Table 2.

**Table 2.** Summary of parameters included in research.

No.	Parameters	Symbol	Unit
1.	Particulate Matter	PM <sub>10</sub>	µg/m <sup>3</sup>
2.	Nitrogen Oxides	NO <sub>x</sub>	ppm
3.	Sulphur Dioxide	SO <sub>2</sub>	ppm
4.	Nitrogen Dioxide	NO <sub>2</sub>	ppm
5.	Ozone	O <sub>3</sub>	ppm
6.	Carbon Monoxide	CO	ppm
7.	Wind Speed	WS	km/hr
8.	Temperature	T	°C
9.	Relative Humidity	RH	%

2.3. Multiple Linear Regression (MLR)

The MLR model represents a function of a number of certain parameters that includes one dependent variable and two or more independent variables used as inputs. The application of MLR has been widely used in relation to meteorological indicators to predict PM<sub>10</sub> concentrations, future ozone concentrations, and other atmospheric pollutants [15]. In this study, MLR model was developed by using eighty percent of the dataset for forecasting the next-day PM<sub>10</sub> (PM<sub>10+24</sub>) concentrations at each study location, while the remaining twenty percent is used for model validation. The relationship between dependent and independent variables in MLR model is expressed in general equation as follows [15]:

$$y = b_0 + \sum_{i=1}^n b_i x_i + \varepsilon$$

where  $b_0$  is the intercept,  $b_i$  are the coefficients for the regression,  $x_i$  are the independent variables and  $\varepsilon$  is stochastic error linked with the regression.

2.4. Quantile Regression (QR)

Besides MLR, QR was used to forecast the next-day PM<sub>10</sub> (PM<sub>10+24</sub>) level at Kelang and Petaling Jaya. QR is a modified median regression that involves approximating the value of the parameter vector  $\beta$  from the range of acceptable vectors that lessens the mean loss function. QR permits the variables to have various distributions at various quantiles of the modelled variable distribution and it is robust to deviations from normality and to skewed tails [16]. This study adopted 3 quantiles at each location

therefore 3 equations are generated the same number of quantile regression coefficients. Eighty percent of the data were used for training while twenty percent of data were used for validation. A linear QR model is expressed in the equation below [16]:

$$Y = \beta_0^{(p)} + \beta_1^{(p)}(X_1) + \beta_2^{(p)}(X_2) + \dots + \beta_n^{(p)}(X_n) + \varepsilon_i^{(p)}$$

where  $Y$  is the dependant variable,  $\beta_0$  implies the intercept,  $\beta_1$  until  $\beta_n$  is the slopes of the covariates and  $\varepsilon_i$  is the error term. The  $p$  indicates the  $p^{th}$  quantile, where the value ranges between 0 and 1.

2.5. Model performance evaluation

The performances of the models developed to forecast the next-day PM<sub>10</sub> concentrations at Kelang and Petaling Jaya were assessed using model performance evaluation. In this research, three performance indicators were utilised to determine the model with best performance in forecasting the next-day PM<sub>10</sub> concentration during HPE. The performance indicators are Normalized Absolute Error (NAE), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The formula of the model performance evaluation mentioned above are given as follows.

**Table 3.** Model performance evaluation.

Performance Indicator	Formula	Description
Root Mean Square Error [10]	$RMSE = \frac{1}{n-1} \sum_{i=1}^n (P_i - O_i)^2$	RMSE value closer to 0 indicates better method
Normalized Absolute Error [10]	$NAE = \frac{\sum_{i=1}^n  P_i - O_i }{\sum_{i=1}^n O_i}$	NAE value closer to 0 indicates better method
Mean Absolute Error [10]	$MAE = \frac{\sum_{i=1}^n  P_i - O_i }{n}$	MAE value closer to 0 indicates better method

Where  $n$  is sum of hourly measurements at exact site,  $P_i$  is predicted values of a set of hourly monitoring record and  $O_i$  is observed values of a set of hourly monitoring record.

**3. Result and discussion**

3.1. Data Summary

Table 4 summarizes the descriptive statistics of PM<sub>10</sub> concentrations at each study area in year 1997, 2005, 2013 and 2015. The unit of measurement is microgram per cubic meter (µg/m<sup>3</sup>). The mean of PM<sub>10</sub> concentrations at both study areas surpassed the stipulated New Malaysia Ambient Air Quality Standard (NMAAQs) for the yearly average 40 µg/m<sup>3</sup>. The highest mean at Kelang and Petaling Jaya were observed in year 1997, with the value 87.41 µg/m<sup>3</sup> and 69.42 µg/m<sup>3</sup>, respectively. Petaling Jaya also showed the highest value of mean in 2015. The lowest mean at both locations were observed in 2013, with the value 64.79 µg/m<sup>3</sup> at Kelang and 48.43 µg/m<sup>3</sup> at Petaling Jaya.

**Table 4.** Descriptive statistics of PM<sub>10</sub> concentrations data.

	Kelang				Petaling Jaya			
	1997	2005	2013	2015	1997	2005	2013	2015

N	Valid	8631	8744	8653	8622	8222	8727	8659	8222
	Missing	129	16	107	138	538	33	101	538
Mean		87.41	78.54	64.79	76.96	69.42	64.32	48.43	69.42
Median		70.00	68.00	58.00	64.00	49.00	56.00	43.00	49.00
Mode		48.00	62.00	60.00	56.00	37.00	56.00	38.00	37.00
Std. Deviation		59.90	53.09	43.00	48.18	55.05	40.70	29.31	55.05
Range		396.00	619.00	573.00	388.00	373.00	474.00	355.00	373.00
Minimum		18.00	24.00	22.00	0.00	20.00	20.00	17.00	20.00
Maximum		414.00	643.00	595.00	388.00	393.00	494.00	372.00	393.00
Percentile	25	46.00	54.00	47.00	52.00	39.00	48.00	35.00	39.00
	50	70.00	68.00	58.00	64.00	49.00	56.00	43.00	49.00
	75	102.00	86.00	70.00	82.00	76.00	68.00	54.00	76.00

3.2. Development of prediction model

3.2.1. Multiple Linear Regression (MLR) model. Table 5 shows the summary of MLR model in predicting the next-day PM<sub>10</sub> (PM<sub>10+24</sub>) concentrations in Kelang and Petaling Jaya. The model equations at each location were established by nine parameters which are PM<sub>10</sub>, WS, T, H, NO<sub>x</sub>, SO<sub>2</sub>, NO<sub>2</sub>, O<sub>3</sub> and CO. This approach showed that the effect of previous PM<sub>10</sub>, WS, T, H, NO<sub>2</sub> and CO were consistent at both locations. The equations indicated that NO<sub>2</sub> and CO showed a positive correlation with PM<sub>10</sub> concentrations meanwhile the weather parameters have a negative correlation with the PM<sub>10</sub> concentrations.

Table 5. Model summary of PM<sub>10</sub> concentration using MLR model.

Location	Models
Kelang	$PM_{10+24} = 163.604 + 0.639PM_{10} - 1.325WS - 2.155T - 1.075H + 0.262NO_x + 0.869SO_2 + 0.087NO_2 - 0.133O_3 + 5.810CO$
Petaling Jaya	$PM_{10+24} = 81.475 + 0.599PM_{10} - 0.675WS - 1.106T - 0.434H - 0.065NO_x - 0.163SO_2 + 0.552NO_2 + 0.147O_3 + 3.867CO$

3.2.2. Quantile Regression (QR) model. Table 6 indicates the summary of QR model in predicting the next-day (PM<sub>10+24</sub>) concentrations in Kelang and Petaling Jaya. QR model with quantiles from 0.3 to 0.9 with increment of 0.3 were developed to forecast the next-day PM<sub>10</sub> concentration. The equations indicated that T and H were consistent at all quantile with a negative correlation with PM<sub>10</sub> concentrations at both Kelang and Petaling Jaya.

Table 6. Model summary of PM<sub>10</sub> concentration using QR model.

Location	Quantile	Models
Kelang	0.3	$PM_{10+24} = 89.062 + 0.428PM_{10} - 0.703WS - 0.915T - 0.468H + 0.265NO_x + 0.076SO_2 - 0.066NO_2 - 0.023O_3 + 1.818CO$
	0.6	$PM_{10+24} = 120.665 + 0.650PM_{10} - 0.829WS - 1.471T - 0.697H + 0.244NO_x + 0.254SO_2 - 0.230NO_2 - 0.076O_3 + 2.964CO$

	0.9	$PM_{10+24} = 138.491 + 1.249PM_{10} - 1.119WS - 1.857T - 0.919H + 0.189NO_x + 0.337SO_2 - 0.393NO_2 - 0.180O_3 + 1.448CO$
Petaling Jaya	0.3	$PM_{10+24} = 56.294 + 0.387PM_{10} - 0.810WS - 0.622T - 0.264H + 0.033NO_x + 0.620SO_2 + 0.683NO_2 - 0.139O_3 + 1.639CO$
	0.6	$PM_{10+24} = 71.254 + 0.621PM_{10} - 0.588WS - 0.921T - 0.329H - 0.140NO_x - 0.068SO_2 + 0.720NO_2 + 0.072O_3 + 0.996CO$
	0.9	$PM_{10+24} = 75.594 + 1.148PM_{10} + 0.004WS - 1.053T - 0.332H + 0.028NO_x - 1.918SO_2 + 0.572NO_2 + 0.591O_3 - 0.782CO$

3.2.3. *Performances of the models.* Table 7 tabulated the results for model evaluation performances in predicting the PM<sub>10</sub> concentration for next day (PM<sub>10+24</sub>) at both study areas. Overall, QR models outperformed the prediction of PM<sub>10+24</sub> for both of the places. The result obtained showed that the MAE and NAE values in Kelang indicated that QR model of 0.3 is the best percentile for PM<sub>10</sub> concentration model in Kelang. QR (0.3) was picked as the best model as the model has fewer amount of outliers compared to the rest of models. In Petaling Jaya, all the three error measures showed that QR model with quantile of 0.6 gave better fit than other models as the value gave the smallest error. Hence, QR (0.3) was chosen as the model with best fit, followed by QR (0.6), MLR and QR (0.9). QR (0.6) was selected as the best quantile in Petaling Jaya, followed by MLR, QR (0.3) and lastly QR (0.9).

**Table 7.** Performance comparison of PM<sub>10</sub> prediction models for Petaling Jaya and Shah Alam.

Study Location	Forecasting Day	Method	MAE	NAE	RSME
Kelang	Next-day	MLR	27.85	0.36	42.85
		QR <sub>(τ=0.3)</sub>	<b>18.98</b>	<b>0.25</b>	31.85
		QR <sub>(τ=0.6)</sub>	21.31	0.28	<b>31.47</b>
		QR <sub>(τ=0.9)</sub>	52.47	0.69	63.65
Petaling Jaya	Next-day	MLR	15.77	0.26	24.29
		QR <sub>(τ=0.3)</sub>	16.95	0.28	30.64
		QR <sub>(τ=0.6)</sub>	<b>15.08</b>	<b>0.25</b>	<b>23.36</b>
		QR <sub>(τ=0.9)</sub>	34.35	0.57	40.44

#### 4. Conclusion

Four models were developed to predict next-day PM<sub>10</sub> air pollutant concentrations at Kelang and Petaling Jaya. The prediction model with the best performance was selected based on the model evaluation performance namely MAE, NAE and RMSE. The QR model at percentile of 0.3 was chosen as the best model in Kelang whereas in Petaling Jaya, the QR model at percentile of 0.6 gave the best prediction. Thus, it can be concluded that the best quantile in Kelang is QR (0.3), followed by QR (0.6), MLR and lastly QR (0.6). Meanwhile in Petaling Jaya, quantile 0.6 gave the best fit followed by MLR, 0.9 and lastly QR (0.9). The findings of this study could be used to aid improvements in environmental management policies to reduce the impact of high particulate event in Kelang Valley, Malaysia. It should be emphasized that only models at two monitoring stations in Klang Valley were compared in this study. This can set a boundary to the performance and comparison of the models. Therefore, additional research using data from various monitoring stations is suggested.

## Acknowledgement

Author would like to thank the Ministry of Higher Education Malaysia for the Fundamental Research Grant Scheme (FRGS/1/2020/TK0/UNIMAP/02/53) and the Department of Environment, Malaysia (DOE) for providing the air quality dataset.

## References

- [1] Tan J, Duan J, Chen D, Wang X, Guo S, Bi X, Sheng G, He K and Fu J 2009 Chemical characteristics of haze during summer and winter in Guangzhou *Atmos. Res.* **94** 238–45
- [2] Kunii O, Kanagawa S, Yajima I, Hisamatsu Y, Yamamura S, Amagai T and Ismail I T S 2002 The 1997 haze disaster in indonesia: Its air quality and health effects *Arch. Environ. Health* **57** 16–22
- [3] Liu Q, Ma T, Olson M R, Liu Y, Zhang T, Wu Y and Schauer J J 2016 Temporal variations of black carbon during haze and non-haze days in Beijing *Sci. Rep.* **6**
- [4] Sun J, Wu F, Hu B, Tang G, Zhang J and Wang Y 2016 VOC characteristics, emissions and contributions to SOA formation during hazy episodes *Atmos. Environ.* **141** 560–70
- [5] Glover D and Jessup T 2006 *Indonesia's Fires and Haze: The Cost of Catastrophe* (Institute of Southeast Asian Studies; International Development Research Centre)
- [6] Abdullah S, Napi N N L M, Ahmed A N, Mansor W N W, Mansor A A, Ismail M, Abdullah A M and Ramly Z T A 2020 Development of multiple linear regression for particulate matter (PM10) forecasting during episodic transboundary haze event in Malaysia *Atmosphere (Basel)*. **11** 1–14
- [7] Abdullah S, Ismail M, Fong S Y and Ahmed A N 2016 Evaluation for Long Term PM10 Concentration Forecasting using Multi Linear Regression (MLR) and Principal Component Regression (PCR) Models *EnvironmentAsia* **9** 101–10
- [8] Abdullah S, Ismail M and Fong S Y 2017 Multiple Linear Regression (MLR) Models for Long Term PM10 Concentration Forecasting During Different Monsoon Season *J. Sustain. Sci. Manag.* **12** 60–9
- [9] Abdullah S, Ismail M, Ahmed A N and Abdullah A M 2019 Forecasting Particulate Matter Concentration Using Linear and Non-Linear Approaches for Air Quality Decision Support *Atmosphere (Basel)*. **10**
- [10] Ul-Saufie A Z, Yahaya A S, Ramli A and Hamid H A 2012 Future PM10 Concentration Prediction Using Quantile Regression Models *IPCBE* **37** 15–9
- [11] Baur D, Saisana M and Schulze N 2004 Modelling the effects of meteorological variables on ozone concentration - A quantile regression approach *Atmos. Environ.* **38** 4689–99
- [12] Shaziyani W N, Ul-Saufie A Z, Ahmat H and Al-Jumeily D 2021 Coupling of quantile regression into boosted regression trees (BRT) technique in forecasting emission model of PM10 concentration *Air Qual. Atmos. Heal.* **14** 1647–63
- [13] Kudryavtsev A A 2009 Using quantile regression for rate-making *Insur. Math. Econ.* **45** 296–304
- [14] Zhao W, Fan S, Guo H, Gao B, Sun J and Chen L 2016 Assessing the impact of local meteorological variables on surface ozone in Hong Kong during 2000–2015 using quantile and multiple line regression models *Atmos. Environ.* **144** 182–93
- [15] Ul-Saufie A Z, Yahaya A S, Ramli N A, Rosaida N and Hamid H A 2013 Future daily PM10 concentrations prediction by combining regression models and feedforward backpropagation models with principle component analysis (PCA) *Atmos. Environ.* **77** 621–30
- [16] Sayegh A S, Munir S and Habeebullah T M 2014 Comparing the performance of statistical models for predicting PM10 concentrations *Aerosol Air Qual. Res.* **14** 653–65