Prediction of PM_{10} Concentration During Haze Event in Malaysia using Quantile Regression Approach

Nur Alis Addiena A Rahim[1,2], Norazian Mohamed Noor[1,2,*], Izzati Amani Mohd Jafri[1,2]

[1]Faculty of Civil Engineering & Technology, Universiti Malaysia Perlis (UniMAP), Perlis, Malaysia
[2]Sustainable Environment Research Group (SERG), Centre of Excellence Geopolymer and Green Technology (CEGeoGTech), Universiti Malaysia Perlis, Jejawi 02600, Perlis, Malaysia

ABSTRACT

In this paper, the investigation on the haze event in Malaysia, which worsened the local air quality was assessed. Particulate matter with an aerodynamic diameter of less than 10 µm (PM_{10}) is the primary indicator of air quality index (API) during haze. PM_{10} is thought to have negative impact on both human health and the environment. Then, the development of quantile regression (QR) model to predict the future PM_{10} levels need to be achieve. To achieve this, the dataset of PM_{10} concentration with gaseous pollutants and weather parameters in Klang and Petaling Jaya from the year of haze event in Malaysia (1997, 2005, 2013 and 2015) were obtained from the Department of Environment (DOE) Malaysia. The QR models were developed at percentile of 0.2, 0.4, 0.6 and 0.8. From the experimental works, QR model at quantile 0.4 and 0.6 was chosen as the best predictive tools for predicting the next day PM_{10} concentration during haze event in Klang and Petaling Jaya, respectively. 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.

Keywords: Particulate matter, haze event, quantile regression, air quality prediction

1. INTRODUCTION

Air pollution has is one of the major environmental issues in Malaysia for the past years. The local air quality has been degraded due to the presence of various source, including haze. Haze is a meteorological condition characterized by atmospheric visibility of less than 10 kilometers, resulting from the presence of suspended solid or liquid particles, smoke, and vapor [1]. The occurrence of haze is associated with both meteorological conditions and anthropogenic activities, including industrial operations, mobile emissions, and biomass burning [2-3]. Particulate matter with an aerodynamic diameter of less than 10 µm or known as PM_{10} is the significant criterion of air quality parameter as haze-related pollutant. Hence, it is crucial to evaluate and continually monitor PM_{10} levels through forecasting models to enhance air quality. Statistical approaches have been widely applied in air quality assessment in Malaysia such as multiple linear regression (MLR) and quantile regression (QR). MLR has been commonly used in forecasting air pollution [4]. Various studies have been undertaken to create prediction models for PM_{10} concentration in the East Coast of peninsular Malaysia. These studies specifically involved the development of MLR models, considering different site classifications and various monsoon seasons to ascertain variations during non-haze periods [5-7]. Nevertheless, MLR has limitations, as it may fail to capture the response to non-central locations of predictor variables and is unable to fully adhere to model assumptions [8].

QR is an approach that has been used in forecasting the PM_{10} level. It is evolving as a comprehensive approach to the statistical analysis of both linear and nonlinear models [9]. QR signifies the non-central location of a distribution, enabling the approach to be more practical and accurate [10]. QR models possess certain advantages over MLR, as they do not depend on specific properties, are independent or only mildly dependent, exhibit robustness to outliers, and are

* Corresponding authors: norazian@unimap.edu.my

distribution-free [11]. QR indicated better performance in predicting the future PM_{10} concentrations in Seberang Perai, Malaysia, by comparing the performance of QR and MLR [8].

The objective of this paper is to assess develop predictive models for PM_{10} concentration using QR approach and compare the predictive performance of MLR and QR models for next-day PM_{10} levels in Klang and Petaling Jaya. The outcomes of this study could offer valuable insights for authorities, aiding in planning and implementing necessary measures to mitigate exposure to air pollution and enhance air quality in the specified locations.

2. EXPERIMENTAL PROCEDURE

2.1 Study Area

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 Air quality monitoring stations

Location	Background	Specific Location
Sekolah Menengah (P) Raja Zarina	Urban	Klang
Sekolah Kebangsaan Bandar Utama	Urban	Petaling Jaya

2.2 Air Quality Dataset

The dataset were acquired from Department of Environment (DOE) Malaysia which comprises of hourly data of particulate matter with less than 10 microns in size (PM_{10}); gaseous pollutants including nitrogen oxides (NO_x), sulphur dioxide (SO_2), nitrogen dioxide (NO_2), ozone (O_3), carbon monoxide (CO); and weather parameters such as wind speed (WS), ambient temperature (T) and humidity (H) recorded throughout the year where Malaysia experienced haze event (1997, 2005, 2013 and 2015).

2.3 Model Development

The dataset obtained were split into two, used for training and validation. 80 percent of the dataset was applied in model development, meanwhile 20 percent of the data were used in model validation.

Multiple linear regression (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. In this study, MLR model was developed to compare the performance of the model with quantile regression (QR) approach.

The concentration of PM_{10} at each study area were modelled using quantile regression (QR). QR is an extended median regression that involves predicting the parameter vector β from a series of suitable vectors that reduces the mean loss function. The linkage between a set of independent variables and set of percentiles of a dependent variable, most typically the median, is modelled using quantile regression. In this study, four percentiles were adopted in developing the prediction model using QR approach, which are 0.2, 0.4, 0.6 and 0.8.

2.4 Performance Assessment

The effectiveness of the models created for predicting next-day PM10 concentrations in Klang and Petaling Jaya was evaluated through model performance measure. In this study, three performance indicators were employed to identify the model with the most accurate forecasting performance during haze. The performance indicators include Normalized Absolute Error (NAE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE).

3. RESULTS AND DISCUSSION

Table 2 indicates the summary of QR model in predicting the next-day (PM_{10+24}) concentrations in Klang and Petaling Jaya. QR model with quantiles from 0.2 to 0.89 with increment of 0.2 were developed to predict the next-day PM_{10} concentration. The equations revealed a consistent negative correlation between temperature and humidity with PM_{10} concentrations at all quantiles in both Klang and Petaling Jaya.

Table 2 Model summary of PM_{10} concentration using QR model

Location	Quantile	Models
Klang	0.2	$65.991 + 0.341PM_{10} - 0.625WS - 0.467T - 0.33H + 0.341NO_X + 0.108SO_2 - 0.001NO_2 + 0.003O_3 + 1.634CO$
	0.4	$90.770 + 0.506PM_{10} - 0.762WS - 0.925T - 0.498H + 0.197NO_X + 0.144SO_2 - 0.052NO_2 + 0.02O_3 + 1.964CO$
	0.6	$120.665 + 0.650PM_{10} - 0.829WS - 1.471T - 0.697H + 0.244NO_X + 0.254SO_2 - 0.230NO_2 - 0.076O_3 + 2.964CO$
	0.8	$126.244 + 0.895PM_{10} - 0.871WS - 1.636T - 0.774H + 0.369NO_X + 0.404SO_2 - 0.146NO_2 - 0.071O_3 + 2.676CO$
Petaling Jaya	0.2	$56.336 + 0.329PM_{10} - 0.834WS - 0.659T - 0.272H + 0.1NO_X + 0.735SO_2 + 0.601NO_2 - 0.22O_3 + 1.73CO$
	0.4	$58.365 + 0.449PM_{10} - 0.771WS - 0.632T - 0.268H - 0.04NO_X + 0.365SO_2 + 0.68NO_2 - 0.089O_3 + 1.57CO$
	0.6	$71.254 + 0.621PM_{10} - 0.588WS - 0.921T - 0.329H - 0.14NO_X - 0.068SO_2 + 0.720NO_2 + 0.072O_3 + 0.996CO$
	0.8	$88.601 + 0.892PM_{10} - 0.457WS - 1.277T - 0.410H - 0.096NO_X - 0.962SO_2 + 0.496NO_2 + 0.275O_3 - 0.205CO$

The performance of the models in predicting the PM_{10} concentration for next day (PM_{10+24}) are shown in Table 3. Overall, QR models performed the best compared to MLR for the prediction of PM_{10+24} for both of the places. The error values indicated that QR model at p = 0.4 is the best model for the prediction of PM_{10} concentration model in Klang. QR (0.4) was chosen as the best model as the model has fewer amount of outliers compared to the rest of models. In Petaling Jaya, the error measures of the QR model exhibited the lowest value 0.6th percentile, indicating a better fit than other models.

Table 3 Performance of PM_{10} prediction models for Klang and Petaling Jaya

Location	Method	MAE	NAE	RSME
Klang	MLR	27.85	0.36	42.85
	$QR_{(r=0.2)}$	22.7	0.3	38.4
	$QR_{(r=0.4)}$	**17.38**	**0.23**	**28.44**
	$QR_{(r=0.6)}$	21.31	0.28	31.47
	$QR_{(r=0.8)}$	32.17	0.42	41.27
Petaling Jaya	MLR	15.77	0.26	24.29
	$QR_{(r=0.2)}$	19.6	0.33	34.34
	$QR_{(r=0.4)}$	15.29	0.25	27.51
	$QR_{(r=0.6)}$	**15.08**	**0.25**	**23.36**
	$QR_{(r=0.8)}$	23.69	0.39	29.93

4. CONCLUSION

QR models at four percentiles i.e 0.2, 0.4, 0.6 and 0.8 were developed to predict next-day PM_{10} concentrations in Klang and Petaling Jaya. The model with the most outstanding performance was chosen by considering error measure metrics, specifically MAE, NAE, and RMSE. The QR model ($p = 0.4$) was chosen as the best model in Klang whereas in Petaling Jaya, the QR model ($p = 0.6$) outperformed the other models. It was proven that QR models outperformed the single MLR model. The results of this study can contribute to the enhancement of environmental management policies aimed at mitigating the impact of elevated particulate events in Malaysia. It is essential to note that this study exclusively compared models at only two monitoring stations in the Klang Valley, which may limit the generalization of performance and comparisons. Consequently, further research utilizing data from diverse monitoring stations is recommended for a more comprehensive understanding.

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] Sun, Y., Zhuang, G., Tang, A., Wang, Y., & An, Z. (2006). Chemical characteristics of PM2.5 and PM10 in haze-fog of PM2.5 and PM10 in haze-fog episodes in Beijing. *Environmental Science and Technology*, 40(10), 3148–3155.

[2] Kunii, O., Kanagawa, S., Yajima, I., Hisamatsu, Y., Yamamura, S., Amagai, T., & Ismail, I. T. S. (2002). The 1997 haze disaster in indonesia: Its air quality and health effects. Archives of Environmental Health, 57(1), 16–22. https://doi.org/10.1080/00039890209602912.

[3] Liu, Q., Ma, T., Olson, M. R., Liu, Y., Zhang, T., Wu, Y., & Schauer, J. J. (2016). Temporal variations of black carbon during haze and non-haze days in Beijing. *Scientific Reports*, 6. https://doi.org/10.1038/srep33331.

[4] Abdullah, S., Napi, N. N. L. M., Ahmed, A. N., Mansor, W. N. W., Mansor, A. A., Ismail, M., … Ramly, Z. T. A. (2020). Development of multiple linear regression for particulate matter (PM10) forecasting during episodic transboundary haze event in Malaysia. *Atmosphere*, 11(3), 1–14.

[5] Abdullah, S., Ismail, M., Ahmed, A. N., & Abdullah, A. M. (2019). Forecasting Particulate Matter Concentration Using Linear and Non-Linear Approaches for Air Quality Decision Support. Atmosphere, 10(667).

[6] Abdullah, S., Ismail, M., & Fong, S. Y. (2017). Multiple Linear Regression (MLR) Models for Long Term PM10 Concentration Forecasting During Different Monsoon Season. *Journal of Sustainability Science and Management*, 12(1), 60–69.

[7] Abdullah, S., Ismail, M., Fong, S. Y., & Ahmed, A. N. (2016). Evaluation for Long Term PM10 Concentration Forecasting using Multi Linear Regression (MLR) and Principal Component Regression (PCR) Models. *Environment Asia*, 9(2), 101–110.

[8] Ul-Saufie, A. Z., Yahaya, A. S., Ramli, A., & Hamid, H. A. (2012). Future PM10 Concentration Prediction Using Quantile Regression Models. IPCBEE, 37, 15–19. Retrieved from http://www.ipcbee.com/vol37/003-ICEAE2012-A00006.pdf.

[9] 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. https://doi.org/10.1016/j.atmosenv.2004.05.028.

[10] Shaziayani, W. N., Ul-Saufie, A. Z., Ahmat, H., & Al-Jumeily, D. (2021). Coupling of quantile regression into boosted regression trees (BRT) technique in forecasting emission model of PM10 concentration. *Air Quality, Atmosphere and Health*, 14(10), 1647–1663. https://doi.org/10.1007/s11869-021-01045-3.

[11] Kudryavtsev, A. A. (2009). Using quantile regression for rate-making. *Insurance: Mathematics and Economics*, 45, 296–304. https://doi.org/10.1016/j.insmatheco.2009.07.010.