E3S Web of Conferences **437**, 01006 (2023) https://doi.org/10.1051/e3sconf/202343701006
IConGEET2023

Short-term Predictions of PM_{10} Using Bayesian Regression Models

Norazrin Ramli [1,2*], *Hazrul* Abdul Hamid [3], *Ahmad Shukri* Yahaya[4], *Norazian Mohamed* Noor [1,2] and *Holba*n Elena[5]

[1] Faculty of Civil Engineering & Technology, Universiti Malaysia Perlis, Jejawi, 02600 Arau, Perlis, Malaysia
[2]Sustainable Environment Research Group (SERG), Centre of Excellence Geopolymer and Green Technology (CEGeoGTech), Universiti Malaysia Perlis, Jejawi, 02600 Arau, Perlis, Malaysia
[3]School of Distance Education, Universiti Sains Malaysia, 11800 Gelugor, Penang, Malaysia; hazrul@usm.my (H.A.H)
[4]School of Civil Engineering, Engineering Campus, Universiti Sains Malaysia, 14300 Nibong Tebal, Pulau Pinang, Malaysia
[5]National Instiute for Research and Development in Environmental Protection, Splaiul Independenţei 294, Bucharest, Romania, 060031

Abstract. One of the air pollutants that poses the greatest threat to human health is PM_{10}. The objectives of this study are to develop a prediction model for PM_{10}. The Multiple Linear Regression (MLR) and Bayesian Regression (BRM) models were constructed to forecast the following day's (Day 1) and next two days' (Day 2) PM_{10} concentration. To choose the optimal model, the performance metrics (NAE, RMSE, PA, IA, and R^2) are applied to each model. Jerantut, Nilai, Shah Alam, and Klang were chosen as monitoring sites. Data from the Department of Environment Malaysia (DOE) was utilised as a case study for five years, with seven parameters (PM_{10}, temperature, relative humidity, NO_2, SO_2, CO, and O_3) chosen. According to the findings, the key factors responsible for the unhealthy levels of air quality at the Klang station include carbon monoxide (CO), nitrogen dioxide (NO_2), sulphur dioxide (SO_2), and ozone (O_3) from industrial and maritime activities, which are thought to influence PM_{10} concentrations in the area. When compared to MLR models, the results demonstrate that BRM are the best model for predicting the next day and next two days PM_{10} concentration at all locations.

1 Introduction

PM is an abbreviation for "particulate matter," which refers to all of the solid and liquid particles that are suspended in the air and include such from dust, pollen, soot, smoke, and liquid droplets. The particles' size and composition may vary[1]. According to a study [2], PM has a variety of physical qualities, including particle size and quantity, total surface area, electrostatic properties, and biological and chemical components. PM is a composite of

* Corresponding author: norazrin@unimap.edu.my

several chemical species, rather than a single pollutant. A national study of metropolitan populations in the United States discovered that short-term exposure to PM components was related with higher mortality [3]. The health effects of PM features were studied, and it was discovered that the toxicity of PM varied depending on particle size, location, and season [4]. The yearly open burning of biomass on Sumatra, Indonesia has worsened the air quality in Malaysia, resulting in a transboundary haze. In Indonesia, forest fires have risen due to El Nino, inappropriate forest management, and lack of fire suppression [5]. PM_{10} concentrations in Asia and the Pacific continue to be the leading cause of air pollution issues [6] considered the most harmful pollution in Peninsular Malaysia and Southeast Asia[7] Prediction model are useful tool for predicting PM_{10} concentrations in Malaysia since they are meant to reduce autocorrelation or inaccuracy in the model. The statistical modelling has the capacity to predict PM_{10} concentrations with high precision [8,9]. In the framework of statistical study, meteorological data and data gathered from monitoring air pollution were used as predictive tools in regression procedures. When compared to nonlinear models, linear models are easier to understand and are more straightforward. A multiple linear regression (MLR) model was used as one of the modelling strategies used to estimate PM_{10} concentrations [10–12]. As a result, this study was conducted to forecast PM_{10} concentrations using a MLR and a Bayesian regression model (BRM).This study's objectives are to identify the descriptive analysis of PM_{10}, meteorology, and gaseous parameters; to predict PM_{10} concentrations for the next day (Day 1) and the next two-days (Day 2) using MLR and BRM; and to determine the best model for predicting PM_{10} concentrations for the next day and the next two days using Performance Indicators.

2 Methods

2.1 Description of study area

The four stations used to forecast PM_{10} concentration were Nilai, Shah Alam, Klang, and Jerantut. In central Peninsular Malaysia, these stations serve as a background station (Jerantut), an industrial area (Nilai), and two urban areas (Shah Alam and Klang).

2.2 Data

Monitoring data were gathered from the Malaysian Department of Environment (DOE) for the period of January 2008 to December 2012. Data analysis was done using multiple linear regression and the Bayesian regression model. Particulate matter smaller than 10 microns is the dependent variable in this study, and the independent variables are temperature, relative humidity, sulphur dioxide, nitrogen dioxide, carbon monoxide, and ozone. Table 1 provides a summary of variable. The data from the air monitoring system was split into two groups: training and validation. 80 percent of all monitoring data as used for training data, and the remaining 20 percent was used for validation.

Table 1. The summary of variable

Symbol	Parameter	Unit	Variable
PM_{10}	Particulate matter	$\mu g/m^3$	Dependent
$PM_{10,D0}$	PM_{10} at day 0	$\mu g/m^3$	Independent
NO_2	Nitrogen dioxide	ppm	Independent
SO_2	Sulphur dioxide	ppm	Independent

CO	Carbon monoxide	ppm	Independent
O_3	Ground- level ozone	ppm	Independent
T	Temperature	°C	Independent
RH	Relative humidity	%	Independent
WS	Wind speed	km/hr	Independent

2.3 Statistical analysis

Multiple linear regression (MLR), the Bayesian regression model with non-informative prior (BRM-NIP), and the Bayesian regression model with conjugate prior are the three regression models employed in this study (BRM-CP).

2.3.1 Multiple linear regression (MLR)

There is a value of the dependent variable Y for each value of the independent variable X. The general equation for a multiple linear regression with independent variables given n observations is shown in Equation (1).

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_7 X_{7i} + \varepsilon_i \quad (1)$$
$$i = 1, 2, \cdots, n$$

In a regression model, β's is treated as a constant while Y and X are treated as variables. By minimizing the sum of the squares of the vertical deviation from the data point to the line, statistical software determines the least square model's best fitting line for the observed data [13]. Temperature (T), relative humidity (RH), nitrogen dioxide (NO_2), sulphur dioxide (SO_2), carbon monoxide (CO), ozone (O_3), and particulate matter at that time (PM_{10D0}) are the parameters used in the model. However, for some cases, such as missing data and limitation of data, the independent parameter is not included in the regression model.

2.3.2 Bayesian regression (BRM)

The posterior distribution P (θ|x) is proportional to the product of the likelihood function P(x|θ) and the prior distribution of the parameter P(θ) [12]. Bayes rule is applied in Bayesian statistics by,

$$\textit{Pr (Posterior distribution) } \alpha \textit{ Pr (Likelihood) x Pr (Prior distribution)} \quad (2)$$

$$\textit{Pr (Posterior distribution) } \alpha \textit{ Pr (gamma) x Pr (normal or uniform)} \quad (3)$$

Beta's (β's) and X are both considered random variables in Bayesian statistics. As a result, the likelihood distribution for x is thought to be normal and the previous distribution for beta is uniform or normal [14]. In this study, the prior distribution used is Beta (β) is uniform and Tau (τ) following a gamma distribution and Beta (β) following normal distribution [15]. Priors are classified into two types: conjugate priors and non-informative priors. A conjugate prior is defined as a normal-gamma prior when the posterior distribution has the same form as the prior distribution. When there is little or no information on the prior, a non-informative prior is utilised[16]. Based on a random sample of 2000 observations, this study inferred [15]. Bayesian analysis often uses the Markov Chain Monte Carlo (MCMC) method. The techniques used to sample the posterior probability distribution are Gibbs sampling and Metropolis-Hasting [17]. Based on the software WinBUGS, the Gibbs sampling methods running from the R package are used to get the posterior distribution of the parameters [13].

2.3.3 Performance indicators (PI)

Calculating the performance indicators allows for an evaluation of the model's performance. Coefficient of determination (R^2), index of agreement (IA), prediction accuracy (PA), normalised absolute error (NAE), and root mean square error (RMSE) have all been employed as performance indicators. The formula for the performance indicator obtained derived from Table 2.

Table 2. The performance indicator

PI	Equation	Criteria
R^2	$R^2 = \left(\frac{\sum_{i=1}^{n}(P_{i} - P)\,(O_{i} - O)}{n.S_{pred}.S_{obs}}\right)$	closer to one (1)
IA	$IA = 1 - \left(\frac{\sum_{i=1}^{n}(P - O_i)^2}{\sum_{i=1}^{n}(\lvert P - \bar{O}\rvert + \lvert O_i - \bar{O}\rvert)^2}\right)$	
PA	$PA = \left(\frac{\sum_{i=1}^{n}(P_{i} - \bar{O})^2}{\sum_{i=1}^{n}(O_{i} - \bar{O})^2}\right)$	
NAE	$NAE = \frac{\sum_{i=1}^{n}\lvert P_{i} - O_i\rvert}{\sum_{i=1}^{n} O_i}$	closer to the smallest values or zero (0).
RMSE	$RMSE = \frac{1}{n-1}\sum_{i=1}^{n}(P_{i} - O_i)^2$	

3 Results and Discussions

3.1 Descriptive analysis for PM_{10}

The result and summary of descriptive analysis is presented in Table 3. The comparison of PM_{10} concentration is referred to the Malaysian Ambient Air Quality Guidelines (MAAQG) interim target-1 (2015) since the datasets is between 2008 to 2012.

Table 3. The descriptive analysis of PM_{10} concentration for 2008-2012

	Jerantut	Shah Alam	Klang	Nilai
Mean	39.80	57.50	70.81	64.64
Median	38.00	54.00	66.00	62.00
SD*	15.11	19.48	26.83	18.83
CV*	0.38	0.34	0.38	0.29
Skewness	0.60	1.60	2.09	1.09
Kurtosis	0.52	7.38	8.28	2.28
Maximum	104.00	212.00	266.00	160.00
Percentile:				
25^{th}	29.00	45.00	54.00	52.00
50^{th}	38.00	54.00	66.00	62.00
75^{th}	49.00	66.00	80.00	74.00

*SD: standard deviation *CV: coefficient of variation

The Klang station has the highest mean PM_{10} concentration (266 g/m^3), followed in decreasing order by Shah Alam (212 g/m^3), Nilai (160 g/m^3), and Jerantut (104 g/m^3). For all locations, the distribution of PM_{10} concentration is biassed to the right. All monitoring stations have positive values for the kurtosis. Klang station (2.09) has the highest skewness value, followed by Shah Alam (1.60) and Nilai (1.09). The skewness score of 0.60 for the PM_{10} concentration at Jerantut station shows a normal distribution. A measurement of the dispersion of the data is the computed coefficient of variance (CV). According to the CV values, Jerantut and Klang monitoring stations have the least variable dispersion of PM_{10} concentration, at 0.38, followed by Shah Alam (0.34) and Nilai (0.29). Jerantut and Klang station exhibit a comparable variance for all PM_{10} concentrations in this investigation. The monitoring station at Nilai had a lower CV, which showed that the data on PM_{10} concentration were dispersed less. This study's analysis reveals that high particulate event happened at three monitoring stations (Shah Alam, Klang, Nilai) between 2008 and 2012. For five years, the average concentration at the Shah Alam, Klang, and Nilai stations was above 50 g/m^3 annually.The PM_{10} concentration at the Jerantut monitoring station is under the interim target-1 of the Malaysian Ambient Air Quality Standard (MAAQS). It is comparable to the outcomes shown by [7] where the concentrations of all pollutants are lower and do not exceed the threshold in Jerantut monitoring station.

3.2 Multiple linear regression model (MLR)

The model's contributors are summarised by the standardised coefficients. The stronger contribution of that variable to the concentration is shown by high values of the standardised coefficient. Table 4 displays the best MLR for PM_{10} for the following day and the following two days for the monitoring stations at Jerantut, Shah Alam, Klang, and Nilai. R^2 values for the Jerantut monitoring station range from 0.519 to 0.803. This demonstrates that the regression model's fitting is good. The main factors for predicting PM_{10} at the Jerantut monitoring station in 2012, according to regression models for the prediction of Day 1 PM_{10} concentration, are temperature, particulate matter on that day (PM_{10D0}) and ozone, which have the highest coefficients values and R^2 values (0.803). The model reveals that the key contributors to PM_{10} concentration on that day (PM_{10D0}), NO_2, and relative humidity, which have the greatest R^2 values, in 2012 at Shah Alam monitoring station are particulate matter (PM_{10}), NO_2, and relative humidity (0.688). Temperature, relative humidity, and particulate matter on that day (PM_{10D0}) are shown in the model to have the highest contributions for forecasting the next day's PM_{10} concentration in 2012 at the Klang monitoring station (0.745). Nilai is close to a significant rail and air transportation network. Based on this research, the model predicts that relative humidity, NO_2, and PM10 at that day (PM_{10D0}) with the greatest R^2 value (0.685) will be the key contributors for predicting the next day's PM_{10} concentration in 2011. The summary of the MLR prediction model is shown in Table 5.

Table 4. The best MLR of the next day and the next two days of PM_{10} at Jerantut, Shah Alam, Klang and Nilai monitoring station

Model	**Beta**	***R*²**	**Beta**	***R*²**
	2012-Jerantut Day1	**0.803**	2012-Jerantut Day2	0.646
Temperature	**0.045**		0.021	
RH	-0.030		-0.016	
SO_2	-Nil-		-Nil-	
NO_2	0.032		-0.020	
O_3	**0.166**		**0.216**	
CO	0.031		-0.073	
$PM_{10,D0}$	**0.726**		**0.690**	

Model	**Beta**	$\boldsymbol{R^2}$	**Beta**	$\boldsymbol{R^2}$
	2012- Shah Alam Day1	**0.688**	2012- Shah Alam Day2	0.436
Temperature	0.028		**0.036**	
RH	**-0.166**		-0.268	
SO_2	0		-0.005	
NO_2	**0.145**		**0.061**	
O_3	0.046		0.030	
CO	-0.011		-0.003	
$PM_{10,D0}$	**0.723**		**0.537**	
	2012-Klang Day1	**0.745**	2012-Klang Day2	0.504
Temperature	**0.513**		**0.730**	
RH	**-0.490**		**-0.702**	
SO_2	-0.028		-0.061	
NO_2	0.052		-0.016	
O_3	0.034		-0.013	
CO	0.086		0.065	
$PM_{10,D0}$	0.698		0.537	
	2011-Nilai Day1	**0.685**	2011-Nilai Day2	0.406
Temperature	0.059		**0.087**	
RH	**-0.085**		**-0.118**	
SO_2	-0.034		-0.057	
NO_2	**0.150**		**0.100**	
O_3	0.003		-0.027	
CO	-0.011		-0.011	
$\mathbf{PM_{10,D0}}$	**0.753**		**0.586**	

Table 5. The summary of prediction model using MLR

Station/ Year	**MLR**
Jerantut 2012	$PM_{10,D1}$ = (1.733) + (0.291 * Temp) - (0.089 * RH) +(109.922 * NO_2) + (195.793 * O_3) +(4.109 * CO)+(0.725 * PM_{10D0}).
	$PM_{10,D2}$ =(6.726)+(0.135 * Temp) - (0.046 * RH) - (67.897 * NO_2) + (254.626 * O_3)- (9.665 * CO) + (0.690 * $PM_{10\ D0}$)
Shah Alam 2012	$PM_{10,D1}$ = (104.430) + (0.063 * Temp) - (1.108 * RH) - (1.879 * SO_2) + (301.810 * NO_2) + (50.268 * O_3) - (0.573 * CO) + (0.723 * PM_{10D0})
	$PM_{10,D2}$ = (185.036) + (0.082 * Temp) - (1.787 * RH) - (22.452 * SO_2) + (126.342 * NO_2) + (33.250 * O_3) - (0.148 * CO) + (0.538 * $PM_{10\ D0}$)
Klang 2012	$PM_{10,D1}$ = (3.002) + (1.749 * Temp) - (0.659 * RH) - (172.872 * SO_2) + (148.129 * NO_2) +(65.169 * O_3) + (3.665 * CO) + (0.698 * PM_{10D0})
	$PM_{10,D2}$ = (27.719) + (2.529 * Temp) - (0.956 * RH) - (375.042 * SO_2) - (46.328 * NO_2) -(24.721 * O_3) + (2.781 * CO) +(0.537 * PM_{10D0}).

Station/ Year	MLR
Nilai 2011	$PM_{10,D1}$ = (23.447) +(0.205 * Temp) - (0.257 * RH) - (161.847 * SO_2) + (361.301 * NO_2) + (3.126 * O_3) - (0.990 * CO) + (0.751 * PM_{10D0}).
	$PM_{10,D2}$ = (46.441) + (0.304 * Temp) - (0.357 * RH) - (267.937 * SO_2) + (239.331 * NO_2) - (30.380 * O_3) - (0.975 * CO) + (0.583 * PM_{10D0}).

3.3 Bayesian regression model (BRM)

The BRM model was proposed in this study to illustrate a certain model parameter assumption. The prior distribution that was employed was uniform, and it followed both the normal and gamma distributions. Priors can be classified into two categories: non-informative priors (NIP) and conjugate priors (CP). The results of the BRM for the concentration of PM_{10} at the monitoring stations at Jerantut, Shah Alam, Klang, and Nilai are shown in Tables 6 and 7, respectively.

Table 6. The best normal and gamma prior distribution for the next day and the next two days of PM_{10} at all monitoring station

Station/ Year		BRM- Normal and Gamma prior distribution (CP)	R^2
Jerantut	2011	$PM_{10,D1}$ = (-42.62) + (0.01872* Temp) + (0.459* RH) + (190.2* SO_2) + (71.25*NO_2) + (88.08*O_3)+(2.824*CO)+(0.8206* PM_{10D0})	0.665
	2012	$PM_{10,D2}$ = (-8.219E-04)-(1.038E-05* Temp)+(1.185E-05* RH) + (190.2* SO_2) +(0.001904*NO_2)+(6.836E-04*O_3)-(1.51E-05*CO)+(1.0* PM_{10D0})	0.703
Shah Alam	2010	$PM_{10,D1}$ = (37.28)+(0.06102* Temp)-(0.2264* RH) + (106.6* SO_2) + (61.13* NO_2) + (72.09*O_3) + (3.6*CO) + (0.4517* PM_{10D0})	0.654
	2008	$PM_{10,D2}$ = (18.44) + (1.702* Temp) - (0.4832* RH) + (28.28* SO_2)+(77.89* NO_2) -(51.93* O_3) - (1.657*CO) + (0.5187* PM_{10D0})	0.885
Klang	2010	$PM_{10,D1}$ = (39.89)-(0.3245* Temp) -(0.1449* RH) + (103.2* SO_2) + (272.8* NO_2) + (128.4* O_3) + (0.3606* CO)+(0.4715* PM_{10D0})	0.765
	2010	$PM_{10,D2}$ = (16.81)+(1.028* Temp)-(0.06598* RH) + (166.4* SO_2) + (222.0* NO_2) + (71.33* O_3)-(1.431* CO)+(0.1509* PM_{10D0})	0.782
Nilai	2010	$PM_{10,D1}$ = (35.03) - (0.692* Temp) - (0.04624* RH) + (16.95* SO_2) + (465.9* NO_2) + (65.84* O_3) - (2.292* CO) + (0.625* PM_{10D0})	0.772
	2010	$PM_{10,D2}$ = (37.35) - (0.3591* Temp) + (0.05814* RH) + (496.0* SO_2) + (227.9* NO_2) + (37.22* O_3) - (2.178* CO) + (0.394* PM_{10D0})	0.797

Table 7. The best uniform and normal prior distribution for the next day and the next two days of PM_{10} at all monitoring station

Station/ Year		BRM- Uniform and Normal prior distribution (NIP)	R^2
Jerantut	2010	$PM_{10,D1}$ = (-14.27)-(0.1093 * Temp) + (0.1851 * RH) +(67.36 * SO_2) + (170.3 * NO_2) + (219.6 * O_3) + (12.54 * CO) + (0.6329 PM_{10D0})	0.370

	2008	$PM_{10,D2}$ = (-104.1)-(0.3475 * Temp) + (1.312 * RH)-(231.5 * SO_2) + (156.5 * NO_2) + (350.7 * O_3)-(11.8 * CO) + (0.5652 * PM_{10D0})	0.604
Shah Alam	2008	$PM_{10,D1}$ = (-7.848) + (1.324 * Temp)-(0.236 * RH) + (56.12 * SO_2) + (231.5 * NO_2)-(41.64 * O_3)-(1.542 * CO) + (0.6739 * PM_{10D0})	0.543
	2008	$PM_{10,D2}$ = (18.48) + (1.697 * Temp)-(0.482 * RH) + (26.84 * SO_2) + (77.94 * NO_2)-(52.09 * O_3-(1.668 * CO) +(0.5197 * PM_{10D1})	0.100
Klang	2008	$PM_{10,D1}$ = (62.68)-(0.4247 * Temp)-(0.4054 * RH) + (252.5 * SO_2) + (112.3 * NO_2) + (28.61 * O_3) + (2.612 * CO) + (0.6504 * PM_{10D0})	0.579
	2010	$PM_{10,D2}$ = (17.89) + (0.9949 * Temp)-(0.06673 * RH) + (165.8 * SO_2) + (222.5 * NO_2) +(72.24 * O_3)-(1.418 * CO) + (0.151 * PM_{10D0})	0.149
Nilai	2009	$PM_{10,D1}$ = (45.4)-(0.4544 * Temp)-(0.3115 * RH) + (631 * SO_2) + (300 * NO_2) + (141.7 * O_3) +(0.4795 * CO) + (0.7171 * PM_{10D0})	0.432
	2009	$PM_{10,D2}$ = (47.79) +(0.585 * Temp)-(0.4272 * RH) + (482.6 * SO_2) + (28.46 * NO_2)-(12.81 * O_3) + (1.186 * CO) + (0.5504 * PM_{10D0})	0.168

3.4 Performance evaluation

To determine a prediction model's practical accuracy, the validity of models is tested. The study of MLR and BRM is based on the statistical comparison of the model findings with the actual PM_{10} concentration. According to Tables 8, 9, 10, and 11, performance metrics are utilised to compare MLR and BRM for PM_{10} models. The coefficient of determination (R^2), index of agreement (IA), prediction accuracy (PA), normalised absolute error (NAE), and root mean square error (RMSE) are the performance indicators used in this study to identify the best model.

Table 8. Performance index for PM_{10} concentration prediction at Jerantut monitoring station, 2008-2012

Performance Indicator	NAE	RMSE	PA	R^2	IA	Best Model
Day 1 (2008)	0.11	4.98	0.41	0.15	0.62	BRM-CP
Day 2 (2008)	0.13	6.05	0.83	0.60	0.91	BRM-NIP
Day 1 (2009)	0.23	5.32	0.35	0.11	0.56	BRM-CP
Day 2 (2009)	0.38	8.13	0.69	0.42	0.70	BRM-CP
Day 1 (2010)	0.15	4.98	0.65	0.37	0.79	MLR
Day 2 (2010)	0.28	7.82	0.76	0.51	0.83	BRM-CP
Day 1 (2011)	0.11	4.66	0.87	0.66	0.93	BRM-CP
Day 2 (2011)	0.20	7.07	0.81	0.58	0.90	BRM-CP
Day 1 (2012)	0.09	2.71	0.61	0.33	0.77	MLR
Day 2 (2012)	0.12	3.33	0.89	0.70	0.93	BRM-CP

Table 9. Performance index for PM_{10} concentration prediction at Shah Alam monitoring station, 2008-2012

Performance Indicator	NAE	RMSE	PA	R^2	IA	Best Model
Day 1 (2008)	0.09	6.23	0.78	0.54	0.87	BRM-CP
Day 2 (2008)	0.06	3.71	0.98	0.86	0.98	BRM-CP

Performance Indicator	NAE	RMSE	PA	R^2	IA	Best Model
Day 1 (2009)	0.13	10.05	0.81	0.59	0.89	BRM-CP
Day 2 (2009)	0.20	13.60	0.77	0.53	0.86	BRM-CP
Day 1 (2010)	0.12	7.14	0.86	0.65	0.92	BRM-CP
Day 2 (2010)	0.14	7.66	0.89	0.71	0.94	BRM-CP
Day 1 (2011)	0.13	7.85	0.85	0.64	0.92	BRM-CP
Day 2 (2011)	0.18	10.85	0.82	0.60	0.89	BRM-CP
Day 1 (2012)	0.15	6.90	0.47	0.20	0.67	BRM-CP
Day 2 (2012)	0.18	8.35	0.71	0.45	0.84	BRM-CP

Table 10. Performance index for PM_{10} concentration prediction at Klang monitoring station, 2008-2012

Performance Indicator	NAE	RMSE	PA	R^2	IA	Best Model
Day 1 (2008)	0.14	10.05	0.81	0.58	0.81	MLR
Day 2 (2008)	0.22	14.54	0.81	0.58	0.83	BRM-CP
Day 1 (2009)	0.16	11.73	0.80	0.56	0.89	BRM-CP
Day 2 (2009)	0.20	15.05	0.78	0.54	0.88	BRM-CP
Day 1 (2010)	0.08	6.69	0.93	0.76	0.96	BRM-CP
Day 2 (2010)	0.11	8.26	0.94	0.78	0.96	BRM-CP
Day 1 (2011)	0.14	8.88	0.86	0.66	0.92	BRM-CP
Day 2 (2011)	0.18	11.58	0.86	0.65	0.91	BRM-CP
Day 1 (2012)	0.11	8.52	0.53	0.25	0.72	BRM-NIP
Day 2 (2012)	0.15	11.87	0.74	0.49	0.86	BRM-CP

Table 11. Performance index for PM_{10} concentration prediction at Nilai monitoring station, 2008-2012

Performance Indicator	NAE	RMSE	PA	R^2	IA	Best Model
Day 1 (2008)	0.20	12.75	0.54	0.26	0.68	BRM-NIP
Day 2 (2008)	0.29	17.38	0.59	0.31	0.73	BRM-CP
Day 1 (2009)	0.10	6.26	0.91	0.74	0.95	BRM-CP
Day 2 (2009)	0.18	9.33	0.89	0.70	0.88	BRM-CP
Day 1 (2010)	0.09	6.49	0.93	0.77	0.96	BRM-CP
Day 2 (2010)	0.12	7.53	0.95	0.80	0.96	BRM-CP
Day 1 (2011)	0.17	9.80	0.83	0.61	0.90	BRM-CP
Day 2 (2011)	0.28	15.76	0.66	0.38	0.75	BRM-CP
Day 1 (2012)	0.13	8.34	0.68	0.41	0.81	MLR BRM-NIP
Day 2 (2012)	0.18	11.04	0.80	0.56	0.86	BRM-CP

The results demonstrate that, in comparison to MLR, the BRM employing non-informative prior with Tau (τ) following a gamma distribution and Beta (β) following a normal distribution is the best model to predict PM_{10} concentration for all sites. In contrast to MLR models, it has been demonstrated that the BRM is the most accurate model for predicting the following day and the following two days' worth of PM_{10} concentration at all locations.

 https://doi.org/10.1051/e3sconf/202343701006

4 Conclusion

The purpose of the study is to identify the descriptive analysis of PM_{10}, meteorology parameters, and gaseous parameters. This study intended to predict the PM_{10} concentrations for the next day and the next two days using MLR and BRM models. The descriptive analysis shows that the highest mean PM_{10} concentration occurred at Klang station followed by Shah Alam, Nilai and Jerantut. The MLR indicates the good model for PM_{10} concentration prediction in Jerantut, Nilai and Klang for certain years. The performance indicator was applied to MLR, BRM-CP, BRM-NIP model to obtain the best model. The results obtained show that the BRM-CP is the best model to predict the next day and the next two-day PM_{10} concentration at all locations in study.

Acknowledgement

The author would like to thank to Department of Environment Malaysia for the air pollutant dataset.

References

[1] Arita A and Costa M 2011 Environmental Agents and Epigenetics *Handbook of Epigenetics: The New Molecular and Medical Genetics* 459–76

[2] Morakinyo O, Mokgobu M, Mukhola M and Hunter R 2016 Health Outcomes of Exposure to Biological and Chemical Components of Inhalable and Respirable Particulate Matter *Int J Environ Res Public Health* **13** 592

[3] Krall J R, Anderson G B, Dominici F, Bell M L and Peng R D 2013 Short-term Exposure to Particulate Matter Constituents and Mortality in a National Study of U.S. Urban Communities *Environ Health Perspect* **121** 1148–53

[4] Mirowsky J, Hickey C, Horton L, Blaustein M, Galdanes K, Peltier R E, Chillrud S, Chen L C, Ross J, Nadas A, Lippmann M and Gordon T 2013 The effect of particle size, location and season on the toxicity of urban and rural particulate matter *Inhal Toxicol* **25** 747–57

[5] Glover D and Jessup T 1999 *Indonesia's fires and haze : the cost of catastrophe* (Institute for Southeast Asian Studies)

[6] Zhou M, Liu Y, Wang L, Kuang X, Xu X and Kan H 2014 Particulate air pollution and mortality in a cohort of Chinese men *Environmental Pollution* **186** 1–6

[7] Latif M T, Dominick D, Ahamad F, Khan M F, Juneng L, Hamzah F M and Nadzir M S M 2014 Long term assessment of air quality from a background station on the Malaysian Peninsula *Science of the Total Environment* **482–483** 336–48

[8] Ul-Saufie A Z, Yahaya A S, Ramli N A and Hamid H A 2015 PM10 Concentrations Short Term Prediction Using Feedforward Backpropagation and General Regression Neural Network in a Sub-urban Area *Journal of Environmental Science and Technology* **8** 59–73

[9] Shahraiyni H T and Sodoudi S 2016 Statistical modeling approaches for pm10 prediction in urban areas; A review of 21st-century studies *Atmosphere (Basel)* **7**

[10] 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**

[11] Fong S Y, Abdullah S and Ismail M 2018 Forecasting of Particulate Matter (PM10) Concentration Based on Gaseous Pollutants and Meteorological Factors For Different Monsoons of Urban Coastal Area in Terengganu *Journal of Sustainability Science and Management Special Issue Number* **5**

[12] Norazrin R, Hamid H A and Yahaya A S 2023 Boosted Regression Tree (BRT) model for PM10 concentrations prediction in Malaysia *IOP Conference Series: Earth and Environmental Science* vol 1135 (Institute of Physics)
[13] Kery M 2010 *Introduction to WinBUGS for Ecologists: Bayesian approach to regression, ANOVA, mixed models and related analyses*
[14] Pires J C M, Martins F G, Sousa S I V, Alvim-Ferraz M C M and Pereira M C 2008 Prediction of the daily mean PM10 concentrations using linear models *Am J Environ Sci* **4** 445–53
[15] Norazrin R., Yahaya A S and Abdul Hamid H 2019 Predicting PM10 concentration using Bayesian regression with Non-Informative Prior and Conjugate Prior Model *Journal of Engineering and Science Research* **3** 59–65
[16] Evans S 2012 *Bayesian Regression Analysis* (University of Louisville)
[17] Liu Y, Guo H, Mao G and Yang P 2008 A Bayesian hierarchical model for urban air quality prediction under uncertainty *Atmos Environ* **42** 8464–9