

Predicting PM₁₀ concentration using Bayesian regression with Non-Informative Prior and Conjugate Prior Model

Norazrin R.^{a,b*}, Ahmad Shukri Yahaya^a, Hazrul Abdul Hamid^c

^aSchool of Civil Engineering, Engineering Campus,
Universiti Sains Malaysia, 14300 Nibong Tebal, Penang, Malaysia
^bSchool of Environmental Engineering, Universiti Malaysia Perlis,
Kompleks Pusat Pengajian Jejawi 3, 02600 Arau, Perlis, Malaysia
^cSchool of Distance Education, Pusat Pengajian Pendidikan Jarak Jauh,
Universiti Sains Malaysia, 11800 Gelugor, Penang, Malaysia

Abstract: The aim of this study is to predict the next day PM₁₀ concentration using Bayesian Regression with non-informative prior and conjugate prior models. The descriptive analysis of PM₁₀, temperature, relative humidity, nitrogen dioxide (NO₂), sulphur dioxide (SO₂), carbon monoxide (CO) and ozone (O₃) are also included. A case study used two-years of air quality monitoring data at three (3) monitoring stations to predict the future PM₁₀ concentration with seven parameters (PM₁₀, temperature, relative humidity, NO₂, SO₂, CO, and O₃). The descriptive analysis showed that the highest mean PM₁₀ concentration occurred at Klang station in 2011 (71.30 µg/m³) followed by 2012 (68.82 µg/m³). The highest mean PM₁₀ concentration was at Nilai in 2012 (68.86 µg/m³) followed by 2011 (66.29 µg/m³) respectively. The results showed that the Bayesian regression model used a conjugate prior with a normal-gamma prior which was a good model to predict the PM₁₀ concentration for most study stations with ($R^2 = 0.67$ at Jerantut station), ($R^2 = 0.61$ at Nilai station) and ($R^2 = 0.66$ at Klang station) respectively compared to a non-informative prior.

Key words: *Bayesian, Regression, Non-informative Prior, Conjugate Prior, PM₁₀*

INTRODUCTION

Air pollution is the presence of an undesirable material in the atmosphere, in quantities large enough to produce harmful effects. The undesirable materials may affect public health and welfare, vegetation, human property, as well as the global environment [1], [2]. Air pollution in cities is a very complex issue, showing strong seasonality and dependence on meteorological factors. The urban particulate matter with aerodynamic diameter less than 10 µm (PM₁₀) is now identified as one of the most dangerous air pollutants to human health because its size range overlaps with the range of respirable particles [3-5]. In the last decade, several studies have been published about the health effects of the particulate matter. A lengthy exposure to the PM₁₀ and PM_{2.5} can affect the respiratory system and cause cardiovascular diseases, while smaller sized particles will affect childhood

morbidity and mortality [6–9]. High level of PM₁₀ concentration that occurs, especially in the urban areas, could induce an impact on health such as causing pulmonary damage when the particles penetrate into the human respiratory system [10].

In Malaysia, the government body that is responsible for monitoring the air quality is the Department of Environment Malaysia (DOE, Malaysia). Sources of air contaminants may be classified as stationary, mobile, or fugitive which can be attributed to sources such as industrial stack emission, transportation activities such as automobile emissions and sources that are beyond control such as wind-blown dust from stockpiles [1]. The study by Ul-Saufie et al. [11] mentioned that there are three major sources of air pollution in Malaysia which are mobile, stationary and open burning. Simulation models are widely used to estimate the potential impact of air pollution, and a number of methods have been developed by various

Corresponding Author: Norazrin R., School of Civil Engineering, Engineering Campus, Universiti Sains Malaysia, 14300 Nibong Tebal, Penang, Malaysia, e-mail: norazrin.msia@gmail.com

researchers. Air quality forecast models have been employed to provide numerical guidance for forecasters to issue timely ozone and particulate matter concentration forecast to human health exposure [12]. Numerous researchers [6, 11, [13] used multiple linear regression (MLR) method as a forecasting tool because it is easy to implement, calculate and use to develop others models such as AERMOD [7], and the artificial neural network (ANN) [14]. The other analysis that should be considered in air pollution estimation is the Bayesian analysis. Bayesian analysis has been used in various scientific disciplines, including archaeology, conservation biology, ecology and meteorology. [15]. The advantage of the Bayesian data analysis is it allows the research goal with the beliefs about the given data as a probability distribution to be computed. Each variable information represents one or several probability distributions, and the probability is the measure of uncertainty; if the uncertainty is higher, then the probability distribution will be wider [15], [16]. We propose that the objective is to predict the air pollution data using the Bayesian analysis because the uncertainty in air pollution prediction is an important issue that needs to be considered. In Malaysia, the Bayesian model is rarely used for estimating the air pollution data. A Bayesian model is introduced in this study for parameter estimation and uncertainty analysis. Thus, this study is carried out to predict the PM₁₀ concentration using the Bayesian regression model.

MATERIALS AND METHODS

Study area

In this study, three (3) monitoring stations were utilised for predicting future PM₁₀ concentration. One (1) station from an urban area (Klang), one (1) station from an industrial area (Nilai) and one (1) background station (Jerantut) are chosen. Klang is in an urban commercial and industrial area in Selangor and close to the Port Klang. This area is also surrounded by emissions from industries, traffic and marine sources. Nilai is chosen for this study because it is located in a densely industrialised area and near to transportation systems such as highway, railways and also Kuala Lumpur International Airport (KLIA). The background area, Jerantut is situated in central Peninsular Malaysia surrounded by natural forest and agricultural areas. The other two stations are expected to be highly polluted due to the rapid development, industrial activities and traffic emission [17].

Study Flowchart

This study is carried out in the following stages:

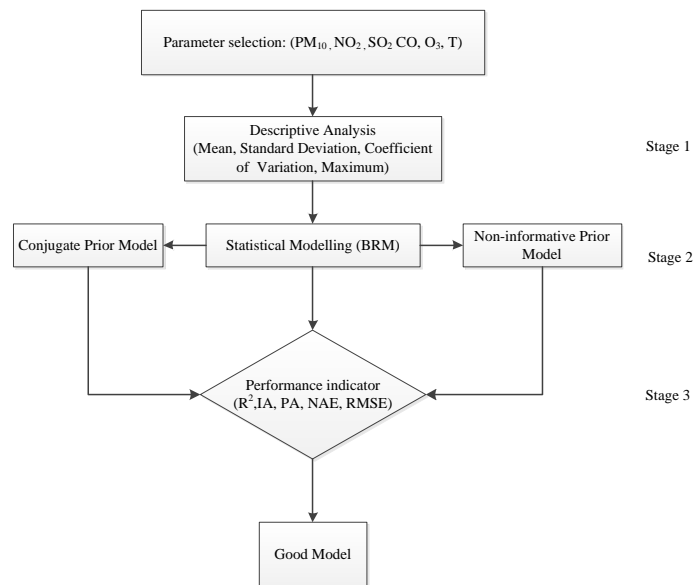


Fig 1 Study Flowchart

Figure 1 represents the study flowchart of this study. The pre-stage in this study is the parameter selection that is used for the BRM model and data screening.

Parameter selection

This study used the daily average monitoring data from January 2011 until December 2012 that was obtained from the Department of Environment (DOE) Malaysia. The data was analysed using a Bayesian regression model with 70 percent (%) of the monitoring data used for model BRM and another 30 percent (%) of the data used for validation. In this study, the dependent variable is a particulate matter less than 10µm (PM₁₀) while the independent variables are PM₁₀ at day 0 (PM₁₀; µg m⁻³), temperature (T; °C), relative humidity (RH; %), nitrogen dioxide (NO₂; ppm), sulphur dioxide (SO₂; ppm), carbon monoxide (CO; ppm) and ozone (O₃; ppm). All these parameters have been found to influence the prediction of PM₁₀ concentrations by other researchers.

Stage 1: Descriptive Analysis

The descriptive analysis is used to determine the characteristics of the monitoring data. Based on that, the central tendency measurements (mean, maximum) and dispersion measurements (standard deviation) are obtained. The descriptive analysis allows for the description of a set of data to be accomplished. There are four general measures of dispersion which are variance, standard deviation, mean and maximum.

Stage 2: Statistical Modelling

Bayesian regression (BRM): The Bayesian approach is used to demonstrate the uncertainty analysis in this study. The basis for the Bayesian inference is Bayes theorem, which is a simple result of conditional probability. The posterior distribution $P(\theta|x)$ is proportional to the product of the likelihood function $P(x|\theta)$ and the prior distribution of the parameter $P(\theta)$ [18]

Bayes rule is applied in the Bayesian statistics by:

$$\text{Pr(Posterior distribution)} \propto \frac{\text{Pr(Likelihood)} \times \text{Pr(Prior distribution)}}{\text{Pr(Prior distribution)}}$$

$$\text{Pr(Posterior distribution)} \propto \frac{\text{Pr(gamma)} \times \text{Pr(normal or uniform)}}{\text{Pr(normal or uniform)}}$$

In Bayesian statistics, beta's (β 's) is considered to be a random variable that is similar to X . So the prior distribution for β is uniform or normal, the likelihood is the distribution for x which is assumed to be normal [18]. In this research, the prior distribution that used β is uniform and Tau (τ) following a gamma distribution and β following a normal distribution. To apply Bayes's theorem, the prior density function is needed where the information is expressed for the unknown parameters [1]. There are two types of priors known as conjugate prior (CP) and non-informative prior (NIP). If no information on the unknown parameters is available, or the information about the prior is negligible or limited [19], the non-informative prior density functions are used which are determined by a constant selected for the vector β of unknown parameters. Conjugate prior (CP) occurs when the posterior distribution has the same form as the prior distribution and defined as normal-gamma prior. Conjugate prior leads after being multiplied by the likelihood function to a posterior distribution which belongs to the same class of the distribution [1]. For this study, CP models that we used are *prior distribution (normal)*, *likelihood (gamma)* and NIP model the *prior distribution (uniform)*, *likelihood (normal)*.

The Markov Chain Monte Carlo (MCMC) method is widely used in the Bayesian analysis. The Gibbs sampling and Metropolis-Hasting algorithms are applied to obtain the sampling for the posterior probability distribution [20]. The posterior distribution of the parameters is obtained using the Gibbs sampling methods. In this study, the practical implementation of MCMC, using Gibbs sampling based on a specialised software WinBUGS [18] running from the R package. The inference of this study is based on a random sample of 2000 observations.

Stage 3: Performance indicators

The model performance is evaluated by calculating the performance indicators. The validity of

the models is performed to estimate the accuracy of a predictive model in practice. Performance indicators that have been used are coefficient of determination (R^2), index of agreement (IA), prediction accuracy (PA) where the best model has the value closer to one (1); while for the normalised absolute error (NAE) and root mean square error (RMSE) the best model has the value closer to zero (0) or the smallest value. The performance indicators equation that used are:

$$R^2 = \left(\frac{\sum_{i=1}^n (P_i - P)(O_i - O)}{n \cdot S_{pred} \cdot S_{obs}} \right)$$

$$IA = 1 - \left(\frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \right)$$

$$PA = \left(\frac{\sum_{i=1}^n (P_i - \bar{O})^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \right)$$

$$NAE = \frac{\sum_{i=1}^n |P_i - O_i|}{\sum_{i=1}^n O_i}$$

$$RMSE = \frac{1}{n-1} \sum_{i=1}^n (P_i - O_i)^2$$

RESULT AND DISCUSSION

The analysis shows that there are several events that occur at two monitoring stations (Klang and Nilai) in 2011 to 2012. The mean concentration for Klang and Nilai stations are above $50 \mu\text{g}/\text{m}^3$ for two years in a row which exceed the Malaysian Ambient Air Quality Guidelines (MAAQG). It is observed that the highest mean PM_{10} concentration occurs at Klang station in 2011 ($71.30 \mu\text{g}/\text{m}^3$) followed by 2012 ($68.82 \mu\text{g}/\text{m}^3$). At Nilai station, the highest mean is shown in 2012 ($68.86 \mu\text{g}/\text{m}^3$) followed by 2011 ($66.29 \mu\text{g}/\text{m}^3$). Based on this study, it is found that Klang and Nilai stations have a high concentration of PM_{10} due to motor vehicles and industrial emissions. These results are in line with the findings of previous studies by Mohamed Noor et al, [21] and Dominick et al, [22] which indicate a similar pattern. According to Ul-Saufie et al. [14], Nilai experiences a large growth in industrial activities which results in higher traffic density and also increase in population. This is consistent with the increase of PM_{10} concentration in that area. Table 3 demonstrates the result of the descriptive analysis of PM_{10} concentration for 2011-2012.

Table 3: The descriptive analysis of PM₁₀ concentration for 2011-2012

Year	2011	2012
Jerantut		
Mean	38.45	37.28
Standard Deviation	14.33	15.33
Coefficient of Variation	0.37	0.41
Maximum	92.00	104.00
Klang		
Mean	71.30	68.82
Standard Deviation	24.15	32.35
Coefficient of Variation	0.34	0.47
Maximum	162.00	266.00
Nilai		
Mean	66.29	68.86
Standard Deviation	20.60	21.44
Coefficient of Variation	0.31	0.31
Maximum	160.00	154.00

At the Jerantut monitoring station, the PM₁₀ concentration does not exceed the Malaysian Ambient Air Quality Guidelines (MAAQG). It is similar to the results demonstrated by Latif et al., (2014) at the Jerantut monitoring station where the concentrations of all the pollutants are lower and do not exceed the guidelines. The CV for all the stations is small which indicates less dispersion of PM₁₀ concentration data for all the monitoring stations. It can also be seen from the results that the Jerantut and Klang stations show a similar variance for all the PM₁₀ concentrations. Nilai monitoring station has a lower CV which indicates a lower dispersion of the PM₁₀ concentration data. These findings are consistent with the findings of past studies by Latif et al. [23], which reveal that during the dry weather, the air quality in Malaysia is influenced by a high number of biomass burning in Southeast Asia which enhances the regional transboundary transportation including in Jerantut which is the air quality background station. Studies by Mohamed Noor et al., 2015 and Juneng et al., 2009 also reveal that a peak concentration of PM₁₀ during the dry season and forest fire incidences during these periods will cause the long-range transportation of particulate across Peninsular Malaysia.

Bayesian Regression Model (BRM): In this study, the BRM model is developed to represent a specific assumption about the model parameters. A prior distribution used is uniformed and following a gamma distribution and normal distribution. There are two types of priors which are conjugate prior (CP) and non-informative prior (NIP), and the results are summarised in Table 4, 5, 6, and 7 respectively.

Table 4: The standardised coefficients for Conjugate prior (CP)

Stations	Jerantut		Nilai		Klang	
	2011	2012	2011	2012	2011	2012
Temperature	-0.78	-1.13	-0.75	0.74	-1.18	0.33
RH	1.23	0.35	0.58	-1.60	0.30	-0.27
SO ₂	-NA-	-NA-	-0.57	-0.50	-1.11	-1.05
NO ₂	-0.53	-0.60	-1.39	-1.65	-1.22	0.74
O ₃	-0.20	-0.23	-1.98	-0.15	-0.98	3.13
CO	0.24	0.38	-0.31	-0.36	-0.34	-0.97
PM _{10,D0}	-0.75	-0.91	-1.19	-0.84	0.16	-0.40
R ²	0.66	0.33	0.61	0.41	0.66	0.25

The independent variables are random variables, and the Bayesian statistics make inference in the entire distribution. The standardised coefficients represent the variables that contribute to the next day PM₁₀ concentration prediction. As illustrated in Table 4, the results show that the main contributor parameters to the next day PM₁₀ concentration in Jerantut are temperature, relative humidity (RH), nitrogen dioxide (NO₂) and also particulate matter on that day (PM_{10,D0}).

The concentration of sulphur dioxide (SO₂) is not indicated in 2011 and 2012 at Jerantut monitoring station. This result is in line with Latif et al., [23] that demonstrate the decrease of SO₂ concentration at the beginning of 2000 may be due to the reduction of Sulphur content in petrol and diesel. At Nilai monitoring station, the higher correlation coefficients are temperature, RH, NO₂, ozone (O₃) and PM_{10,D0}. Meanwhile, sulphur dioxide (SO₂) and nitrogen dioxide (NO₂) have been discovered as the most influential parameters at the Klang monitoring station followed by ozone and carbon monoxide (CO). These results are consistent with other research [24] which found that, at Klang Valley, the vehicular emissions such as sulfur dioxide (SO₂), nitrogen dioxide (NO₂), carbon monoxide (CO) and particulate matter are accounted for more than 70% of the total emissions in the urban areas.

Table 5: Performance Indicator for Conjugate prior (CP)

Station s	Jerantut		Nilai		Klang	
	2011	2012	2011	2012	2011	2012
IA	0.93*	0.77	0.90*	0.81	0.92*	0.72
PA	0.87*	0.61	0.83*	0.68	0.86*	0.53
R ²	0.67*	0.33	0.61*	0.41	0.66*	0.25
RMSE	4.66	2.71	9.80	8.35	8.88	8.53
NAE	0.11	0.09*	0.17	0.13*	0.14	0.11*

*The result is closer to the criteria of Coefficient of Determination (R²), Index of Agreement (IA), Prediction Accuracy (PA), Normalized Absolute Error (NAE), Root Mean Square Error (RMSE)

The analysis of BRM is based on the statistical comparison of the models' results with the actual PM₁₀ emissions. Further analysis shows that the best model for CP is in 2011 at all the monitoring station locations (refer Table 5). So far, there are limited studies on predicting the air pollution using the Bayesian methods in Malaysia, so these results have not been previously described. The standardised coefficients for the NIP distribution are presented in Table 6, and the performance indicators are shown in Table 7 respectively.

Table 6: The standardised coefficients for Non-informative prior (NIP)

Stations	Jerantut		Nilai		Klang	
	2011	2012	2011	2012	2011	2012
Temperature	-1.11	-1.13	-0.49	0.74	-1.18	0.33
RH	1.23	0.35	-2.10	-1.60	0.30	-0.27
SO ₂	-NA-	-NA-	-0.57	-0.50	-1.11	-1.05
NO ₂	-0.53	-0.60	0.98	-1.65	-1.22	0.74
O ₃	-1.84	-0.23	-1.98	-0.15	-0.98	3.13
CO	-0.52	0.38	-0.31	-0.36	-0.34	-0.97
PM _{10,D0}	-1.02	-0.91	-0.75	-0.84	0.16	-0.40
R ²	0.27	0.33	0.37	0.41	0.26	0.25

Compared to the CP distribution, the results of this study did not show a significant difference between the CP and NIP distributions. However, at the Jerantut monitoring station, the NIP distribution indicates that nitrogen dioxide (NO₂), ozone (O₃), carbon monoxide (CO) and particulate matter are the major contributors to predict the next day PM₁₀ concentration. The results also reveal that the more significant contributions for the next day PM₁₀ concentration at most study areas are the temperature, nitrogen dioxide (NO₂), sulphur dioxide (SO₂), carbon monoxide (CO), ozone (O₃) and particulate matter (PM₁₀) of that day (D₀). Temperature and relative humidity show significant influence to the prediction model, and these results support a study by Afroz et al., 2003 which shows that during the dry season, the atmospheric condition is more stable and this would allow the continuous emissions to build up a high concentration. This might be due to the combined effects of the geographical vicinity, wind flow and possibly some trappings of biomass burning plume. The performance indicators that present the best model to fulfil the PI criteria is shown in Table 7. At the Jerantut monitoring station, the results obtained for IA (0.722 to 0.814), PA (0.537 to 0.681) and NAE (0.094 to 0.111) are closer to the criteria. Based on the results obtained, the good model for NIP distribution is in 2012 at all the monitoring stations (Jerantut, Nilai and Klang) with higher significant values.

Table 7: Performance Indicator for Non-informative prior (NIP)

Stations	Jerantut		Nilai		Klang	
	2011	2012	2011	2012	2011	2012
IA	0.72	0.77*	0.78	0.81*	0.71	0.72*
PA	0.55	0.61*	0.64	0.68*	0.54*	0.53
R ²	0.27	0.33	0.37	0.41	0.26	0.25
RMSE	4.82	2.71	10.1	8.34	9.15	8.52
NAE	0.11	0.09*	0.17	0.13*	0.14	0.11*

* The result is closer to the criteria of Coefficient of Determination (R²), Index of Agreement (IA), Prediction Accuracy (PA), Normalized Absolute Error (NAE), Root Mean Square Error (RMSE)

Further analysis is carried out to determine the best model for BRM. The validation for two types of priors CP versus NIP is summarised in Table 8.

Table 8: Performance Indicator for Conjugate prior (CP) and Non-informative prior (NIP)

Stations	Jerantut		Nilai		Klang	
	CP	NIP	CP	NIP	CP	NIP
IA	0.93*	0.77	0.90*	0.81	0.92*	0.72
PA	0.87*	0.61	0.83*	0.68	0.86*	0.53
R ²	0.67*	0.33	0.61*	0.41	0.66*	0.25
RMSE	4.66	2.71	9.80	8.34	8.88	8.52
NAE	0.11	0.09*	0.17	0.13*	0.14	0.11*

* The best models of Conjugate prior (CP) and Non-informative prior (NIP)

As shown in Table 8, the CP distribution is significantly more than the NIP distribution. For this study, the BRM model using CP distribution is good where the PI results show for IA (0.900 to 0.929), PA(0.830 to 0.866) and R² (0.614 to 0.665). They are closer to one (1) which indicate the predictive model with the actual PM₁₀ emissions relatively accurate for all the monitoring stations. Overall, these results indicate that the conjugate prior (CP) has fulfilled the assumptions and the good model for BRM for all of the study stations (Jerantut, Nilai and Klang).

In summary, the analysis for BRM model's validity is performed, and the results indicate that the Bayesian regression using a conjugate prior (CP) with normal-gamma prior is a better model to predict PM₁₀ concentration for most study stations compared to Bayesian regression using Non-informative prior (NIP). Overall, these results are in agreement with Liu et al., (2008) findings which show that the Bayesian model is useful to predict the urban air quality and can be predicted from related contributing factors. The Bayesian regression model is useful for further study to predict the parameter concentration distribution and the air quality status. A further study with more focus on other meteorological parameters such as wind and precipitation are therefore suggested.

CONCLUSION

The particulate matter with an aerodynamic diameter of less than 10 μm (PM_{10}) is one of the most dangerous air pollutants to human health. This study is carried out to predict the PM_{10} concentrations for the next day using the Bayesian Regression (BRM) model. The Performance Indicators are used to determine a good model for predicting PM_{10} concentrations. The descriptive analysis shows that the highest mean PM_{10} concentration occurs at the Klang station followed by Nilai and Jerantut. Based on the CV values, the dispersion of PM_{10} concentration is less varied for most of the study stations. The Bayesian regression (BRM) models are developed to predict the next day PM_{10} concentration using a conjugate prior (CP) with normal-gamma prior and Non-informative prior (NIP). The assumption for the residuals has been checked and found to have fulfilled the assumption. A good NIP model is indicated in 2012 at most study stations while the good model for CP is indicated in 2011 at most study stations (Jerantut, Nilai and Klang). The results indicate that the Bayesian regression using a conjugate prior (CP) with normal-gamma prior is a good model to predict PM_{10} concentration. The performance indicators (NAE, RMSE, PA, IA and R^2) were applied to each model to obtain the good model. The results obtained showed that the Bayesian regression model (BRM) using a conjugate prior (CP) is the most appropriate model to predict PM_{10} concentration. The model can be used as an early warning system that can be used by local authorities or related government agencies to predict air quality statistics.

ACKNOWLEDGMENTS

This study was supported by Ministry of Higher Education Malaysia (MOHE) and Universiti Malaysia Perlis (UniMAP). We would like to thank the Department of Environment Malaysia (DOE) for providing air monitoring data.

REFERENCES

- [1] R. A. Corbitt, *Air Quality Control: Standard Handbook of Environmental Engineering*. McGraw-Hill, 2004.
- [2] S. R. Jose, *Measuring and Modelling Investigation of Environmental Processes*, Volume 2. Boston: WitPress, 1999.
- [3] S. Chiou and C. Tsai, "Measurement of emission factor of road dust in a wind tunnel," *Powder Technol.*, vol. 118, no. 1–2, pp. 10–15, 2001.
- [4] T. M. Zobeck and R. S. Van Pelt, "Wind-induced dust generation and transport mechanics on a bare agricultural field," *J. Hazard. Mater.*, vol. 132, no. 1, pp. 26–38, 2006.
- [5] B. Jancsek-turóczi, A. Hoffer, and I. Nyír, "Sampling and characterization of resuspended and respirable road dust," vol. 65, pp. 69–76, 2013.
- [6] J. C. M. Pires, F. G. Martins, S. I. V. Sousa, M. C. M. Alvim-Ferraz, and M. C. Pereira, "Prediction of the Daily Mean PM_{10} Concentrations Using Linear Models," *Am. J. Environ. Sci.*, vol. 4, no. 5, pp. 445–453, 2008.
- [7] B. Vijay Bhaskar, R. V. Jeba Rajasekhar, P. Muthusubramanian, and A. P. Kesarkar, "Measurement and modeling of respirable particulate (PM_{10}) and lead pollution over Madurai, India," *Air Qual. Atmos. Heal.*, vol. 1, no. 1, pp. 45–55, 2008.
- [8] M. S. Callén, J. M. López, and a. M. Mastral, "Seasonal variation of benzo(a)pyrene in the Spanish airborne PM_{10} . Multivariate linear regression model applied to estimate BaP concentrations," *J. Hazard. Mater.*, vol. 180, no. 1–3, pp. 648–655, 2010.
- [9] T. Kanabkaew, "Prediction of Hourly Particulate Matter Concentrations in Chiangmai, Thailand Using MODIS Aerosol Optical Depth and Ground-Based Meteorological Data," *EnvironmentAsia*, vol. 6, no. 2, pp. 65–70, 2013.
- [10] B. Vijay Bhaskar, R. V. Jeba Rajasekhar, P. Muthusubramanian, and A. P. Kesarkar, "Measurement and modeling of respirable particulate (PM_{10}) and lead pollution over Madurai, India," *Air Qual. Atmos. Heal.*, vol. 1, no. 1, pp. 45–55, 2008.
- [11] A. Z. Ul-Saufie, A. S. Yahaya, N. Ramli, and H. A. Hamid, "Performance of Multiple Linear Regression Model for Long-term PM_{10} Concentration Prediction Based on Gaseous and Meteorological Parameters," *J. Appl. Sci.*, vol. 12, no. 14, pp. 1488–1494, Dec. 2012.
- [12] and R. S. T. Steyn, D.G., *Air Pollution Modelling and its Application XX*. Netherland: Springer, 2010.
- [13] P. Perez, "Combined model for PM_{10} forecasting in a large city," *Atmos. Environ.*, vol. 60, pp. 271–276, 2012.
- [14] A. Z. Ul-Saufie, A. S. Yahaya, N. A. Ramli, N. Rosaida, and H. A. Hamid, "Future daily PM_{10} concentrations prediction by combining regression models and feedforward backpropagation models with principle component analysis (PCA)," *Atmos. Environ.*, vol. 77, pp. 621–630, 2013.
- [15] J. K. Kruschke, "Bayesian data analysis," *Wiley Interdiscip. Rev. Cogn. Sci.*, vol. 1, no. 5, pp. 658–676, 2010.

- [16] L. Uusitalo, "Advantages and challenges of Bayesian networks in environmental modelling," *Ecol. Modell.*, vol. 203, no. 3–4, pp. 312–318, 2007.
- [17] S. Z. Azmi, M. T. Latif, A. S. Ismail, L. Juneng, and A. A. Jemain, "Trend and status of air quality at three different monitoring stations in the Klang Valley, Malaysia," *Air Qual. Atmos. Heal.*, vol. 3, no. 1, pp. 53–64, 2010.
- [18] M. Kery, *Introduction to WinBUGS for Ecologists: A Bayesian approach to regression, ANOVA, mixed models and related analyses*, 1st editio. Elsevier Inc., 2010.
- [19] S. Evans, S. Evans, and D. Brown, "Submitted by," no. May, 2012.
- [20] Y. Liu, H. Guo, G. Mao, and P. Yang, "A Bayesian hierarchical model for urban air quality prediction under uncertainty," *Atmos. Environ.*, vol. 42, no. 36, pp. 8464–8469, 2008.
- [21] N. Mohamed Noor, A. S. Yahaya, M. Abdullah, and Bakri, "Variation of air pollutant (particulate matter - PM10) in peninsular Malaysia : Study in the southwest coast of peninsular Malaysia Variation of Air Pollutant (Particulate Matter - PM 10) in Peninsular Malaysia Study in the southwest coast of peninsul," no. August 2016, 2015.
- [22] D. Dominick, M. Talib, S. M. Zain, and A. Zaharin, "Spatial assessment of air quality patterns in Malaysia using multivariate analysis," vol. 60, pp. 172–181, 2012.
- [23] M. T. Latif, D. Dominick, F. Ahamad, M. F. Khan, L. Juneng, F. M. Hamzah, and M. S. M. Nadzir, "Long term assessment of air quality from a background station on the Malaysian Peninsula," *Sci. Total Environ.*, vol. 482–483, no. 2, pp. 336–348, 2014.
- [24] R. Afroz, M. N. Hassan, and N. A. Ibrahim, "Review of air pollution and health impacts in Malaysia," *Environ. Res.*, vol. 92, no. 2, pp. 71–77, 2003.