PAPER • **OPEN ACCESS**

A Review of PM10 Concentrations Modelling in Malaysia

To cite this article: Wan Nur Shaziayani *et al* 2020 *IOP Conf. Ser.: Earth Environ. Sci.* **616** 012008

IOP Conf. Series: Earth and Environmental Science **616** (2020) 012008 doi:10.1088/1755-1315/616/1/012008

A Review of PM_{10} Concentrations Modelling in Malaysia

Wan Nur Shaziayani[1], Ahmad Zia Ul-Saufie[1], Zuraira Libasin[1], Fuziatul Norsyiha Ahmad Shukri[1], Sharifah Sarimah Syed Abdullah[1], Norazian Mohamed Noor[2,3]

[1]Faculty of Science Computer and Mathematics, Universiti Teknologi Mara, Cawangan Pulau Pinang, Malaysia
[2]School of Environmental Engineering, Universiti Malaysia Perlis, Kompleks Pengajian Jejawi 3, 02600 Arau, Perlis, Malaysia
[3]Sustainable Environment Research Group, Center of Excellence Geopolymer and Green Technology (CEGeoGTech), Universiti Malaysia Perlis, Kompleks Pengajian Jejawi 2, 02600 Arau, Perlis, Malaysia

E-mail: ahmadzia101@uitm.edu.my

Abstract. The purpose of predictive modelling is to predict the variable of interest with reasonable precision, and often to assess the contribution of the independent variables to the dependent variable. In this paper, all of the works examined are aimed at predicting concentrations of outdoor PM_{10} concentrations. The vast majority of the works reported used almost exclusively predictors of the meteorological and source emissions. However, the use of the Hybrid model in predicting PM_{10} concentrations is still not widely used in Malaysia.

1. Introduction

Air pollution occurs when particulate matter and harmful gases in the atmosphere have high enough concentrations to produce harmful effects on human health, plants, human resources, and even the global environment [1]. Air pollution has become a serious environmental issue in Malaysia as a result of open biomass, motor vehicles and industry [2]. In addition, [3,4] stated that biomass burning from wild fires in Indonesia has also been transported to Malaysia especially during the dry season and the southwest monsoon. According to [5], Particulate Matter (PM_{10}) is a pollutant that has the highest air pollutant index (API) value compared to other pollutants in Peninsular Malaysia. Therefore, researcher around the Malaysia had done many studies in order to predict and forecast of PM_{10} concentrations in Malaysia.

The objective of this paper is in a way to review different methods for PM_{10} concentrations prediction in Malaysia. Statistical method is a popular method that used by many researchers to predict PM_{10} concentrations such as Multiple Linear Regression (MLR), Bayesian Autoregressive, time series and Markov Chain. There is a growing interest in recent years on the use of machine learning in predicting and forecasting ambient air pollution. Machine learning have been successfully implemented in many short-term and long-term forecasting applications. The common machine learning method in predicting PM_{10} concentrations are ultilayer Perceptron (MLP), and Artificial

Neural Network (ANN). This study tries to review the methods, which uses various models for forecasting PM_{10} concentrations in Malaysia.

2. Research Methodology

The database i.e., (TITLE-ABS-KEY ("prediction" AND "PM10" AND "Malaysia")) was used to find the results of all documents in Scopus that related to the PM_{10} concentrations modelling in Malaysia. From figure 1, it shows that study for prediction PM_{10} concentrations in Malaysia is highest at 2018, which are 10 documents. For 2019 until 2020, the number of paper published is drop again.

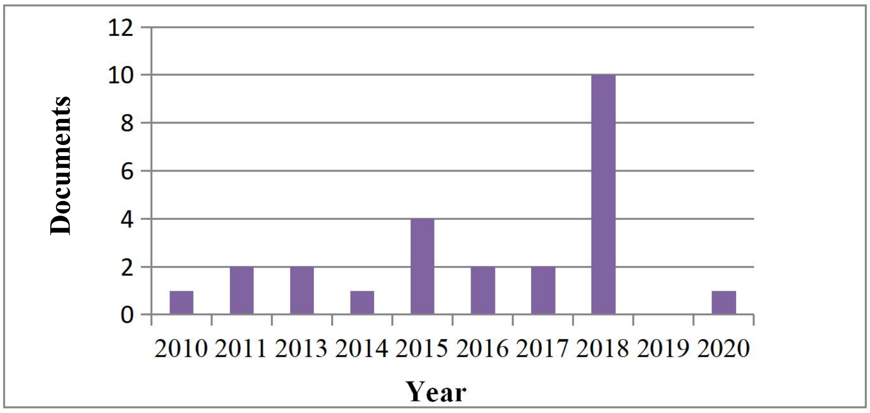

Figure 1. The progress of air pollution models for PM_{10} concentrations prediction.

After reviewing and analysing, figure 2 shows that 72% papers dealing with the use of statistical methods in predicting and forecasting PM_{10} concentrations in Malaysia, 12% used machine learning while another 16% used hybrid model which is combination of machine learning and statistical method.

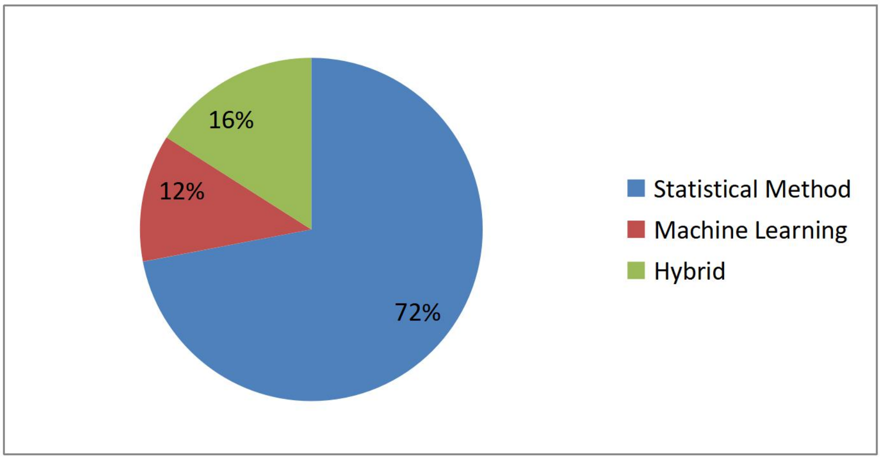

Figure 2. The progress of air pollution models for PM_{10} concentrations prediction.

3. PM_{10} concentrations modelling

The key details of the selected papers, e.g. the authors, year of publication, method and accuracy are presented in this section.

3.1. Statistical method

Table 1 lists the statistical methods in predicting PM_{10} concentrations in Malaysia and the performance indicator that has been used to evaluate and validate the model.

Table 1. The statistical methods in predicting PM_{10} concentrations in Malaysia.

Authors	Year	Method	Accuracy
[6]	2020	MLR	RMSE=126.73-164.98 NAE=0.33-0.43 PA=0.36-0.67
[7]	2018	Bayesian Autoregressive	RMSE=7.56-15.18
[8]	2018	MLR	RMSE=10.66-15.72 R^2=0.57-0.72 NAE=0.14-0.20
[9]	2018	MLR	RMSE=21.63 NAE=0.21 R^2=0.61 IA=0.87 PA=0.78
[10]	2018	Time Series	N/A
[11]	2018	Bayesian Approach	RMSE=3.37-5.3x10^9 MAE=0.68-7.75x10^7 NAE=0.01-1.06x10^6 R^2=0.02-0.99 IA=0.00-0.99 PA=0.13-0.99
[12]	2018	MLR	RMSE=4.08-11.43 MAE=3.23-9.50 MAPE=7.67-24.74
[12]	2018	Regression Time Series Error (RTSE)	RMSE=4.47-10.70 MAE=4.04-8.05 MAPE=7.19-19.17
[13]	2018	Markov Chain	N/A
[14]	2018	MLR	R^2=0.50-0.60
[14]	2018	Stepwise Regression	R^2=0.23-0.29
[15]	2018	MLR	RMSE=11.39-35.71 MAE=8.68-32.75 NAE=0.20-0.79 R^2=0.45-0.72 IA=0.39-0.60
[16]	2017	Dispersion Method	N/A
[17]	2017	Markov Chain	N/A
[18]	2016	Seasonal Time Series	RMSE=11.99 NAE=0.89 MAPE=10.21
[19]	2016	Stepwise Regression	RMSE=6.41-10.87 IA=8.68-32.75 NAE=0.13-0.14 MAE=4.38-8.01

[19]	2016	Time Series (Persistence)	RMSE=6.70-11.68 IA=0.72-0.73 NAE=0.13-0.14 MAE=4.63-8.68
[20]	(2015)	Extreme Value Distributions (EVD)	RMSE=1.75-3233.06 MAE=0.71-385.28 NAE=0.01-5.80 R^2=0.06-0.99 IA=0.00-0.99
[21]	(2015)	Nonlinear Dynamic Concept	N/A
[22]	(2015)	Extreme Value Distributions (EVD)	RMSE=5.01-39.08 MAE=0.95-27.79 NAE=0.02-0.37 R^2=0.88-0.95 IA=0.83-0.99
[23]	2013	MLR	RMSE=14.23-23.03 NAE=0.17-0.23 R^2=0.35-0.62 IA=0.59-0.87 PA=0.59-0.79
[24]	2013	Log Normal Distribution	RMSE=1.47-12.73 NAE=0.01-0.27 R^2=0.24-0.99 IA=0.70=0.99 PA=0.97-0.99
[25]	2011	Probability Distributions (Weibull, log normal, gamma, Rayleigh, Gumbel and Frechet)	RMSE=1.06-133.26 MAE=0.55-25.06 R^2=0.52-0.99 PA=0.73-0.99
[26]	2011	Probability Distributions (Weibull and log normal)	RMSE=2.02-7.27 R^2=0.90-0.99 NAE=0.01-0.06
[27]	2010	Probability Distributions (Weibull and log normal)	RMSE=0.00-15.43 R^2=0.94-0.99 IA=0.97-0.99

From table 1, regression models that are MLR and stepwise regression have been used numerously and successfully in predicting PM_{10} concentrations. This is due to the many statistical software provided this method. There are two types of probability distributions, which are probability density and Extreme value distributions (EVD). The two most used probability density functions are the log normal and Weibull distribution. According to [25,26] Weibull distribution is better than log normal distribution because Weibull gives better accuracy than log normal distribution when the increasing trend of PM_{10} concentration. [20,22] discussed about the EVD and the result showed that the generalized extreme value distribution (GEV) provides better performance indicators in estimating the future extreme concentrations of PM_{10} in Malaysia. We cannot compare the result with [10,11] because it only predict the PM_{10} occurrence.

3.2. Machine learning

Table 2 lists the machine learning methods in predicting PM_{10} concentrations in Malaysia and the performance indicator that has been used to evaluate and validate the model.

Table 2. The machine learning methods in predicting PM_{10} concentrations in Malaysia.

Authors	Year	Method	Accuracy
[28]	2018	ANN	RMSE=12.99 R^2=0.41
[29]	2015	Feedforward Backpropagation Neural Network (FBNN)	R^2=0.88-0.95 IA=0.60-0.85 NAE=0.17-0.22 RMSE=12.81-18.82
[29]	2015	General Regression Neural Network (GRNN)	R^2=0.25-0.50 IA=0.55-0.80 NAE=0.18-0.22 RMSE=13.27-18.94
[30]	2014	ANN	MSE=0.04-0.05 MAE=0.14-0.15

From table 2, [28] showed that the ANN model resulted in a significantly better accuracy compared to MLR. [29] compared the 2001-2010 performance for Seberang Jaya, Pulau Pinang, Malaysia based on gaseous and meteorological parameters and found that the FFBP ANN performs better than the GRNN ANN model. Using ANN thus significantly improved the fitting model with good accuracy.

3.3. *Hybrid Model*

Table 3 lists the combination methods between statistical method and machine learning in predicting PM_{10} concentrations in Malaysia and the performance indicator that has been used to evaluate and validate the model.

Table 3. The combination methods in predicting PM_{10} concentrations in Malaysia.

Authors	Year	Method	Accuracy
[8]	2018	Principle Component Regression (PCR)	RMSE=11.65-19.56 R^2=0.39-0.66 NAE=0.16-0.24
[14]	2018	Principle Component Regression (PCR)	R^2=0.66-0.89
[15]	2018	Principle Component Regression (PCR)	RMSE=11.39-35.71 MAE=8.68-32.75 NAE=0.20-0.79 R^2=0.45-0.72 IA=0.39-0.60
[23]	2013	Principle Component Regression (PCR)	RMSE=11.34-18.27 NAE=0.14-0.19 R^2=0.60-0.76 IA=0.81-0.93 PA=0.78-0.87

From table 3, hybrid models apply the concept of multiple modeling approaches to improve a fit's overall accuracy. Principle Component Regression (PCR) combines the principle component analysis (PCA) and Multiple Linear Regression (MLR) concepts.

4. Conclusions

Due to the current environmental situation, different predictive models and approaches were introduced. This paper addresses the methodological methods, which can be applied to predict for PM_{10} concentrations. Because of the complexities of the system, a single model cannot create the best solution for all situations. Therefore, the most appropriate model depends on the field of analysis and the variables.

Acknowledgement

The authors would like to extend their appreciation to the Department of Environmental Malaysia (DoE) for providing air quality data for this research. The research was funded by 600-IRMI/FRGS 5/3 (289/2019).

References

[1] Norazrin R, Yahaya A S, Abdul Hamid H, Shukri, A, and Abdul H 2018 *J Engineering and Science Research* **2**(2) 59
[2] Kamarul Zaman N A F, Kanniah K D, and Kaskaoutis D G 2017 J Atmospheric Research **193** 142
[3] Kanniah K D, Lim H Q, Kaskaoutis D G, and Cracknell A P 2014 *J. Atmos. Res.* **138**(1) 223
[4] Khan M F, Sulong N A, Latif, M T, Nadzir, M S M, Amil N, Hussain D F M, Lee V, Hosaini P N, Shaharom S, Yusoff N A Y M, Hoque, H M S, Chung J X, Sahani M, Tahir N M, Juneng L, Maulud K N A, Abdullah S M S, Fujii Y, Tohno S, Mizohata A 2016 *J.* Geophys. Res. Atmos. **121**(14) 589–14 611
[5] Fong S Y, Abdullah A and Ismail M 2018 *J. Sustainability Science and Management* **74**(5) 637
[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 *J. Atmosphere* **11**(3) 1
[7] Manga E and Awang N 2018 *J. Stochastic Environmental Research and Risk Assessment* **32**(12) 3409
[8] Ismail M, Abdullah S, Jaafar A D, Ibrahim T A E, and Shukor M S M 2018 *AIP Conference Proceedings,* **2020**
[9] Shaziayani W N, Ul-Saufie A Z and Ahmat H 2018 *AIP Conference Proceedings,* **2013**
[10] AL-Dhurafi N A, Masseran N and Zamzuri Z H 2018 *Stochastic Environmental Research and Risk Assessment* **32**(10) 2903
[11] Ahmat H and Yahaya A S 2018 *AIP Conference Proceedings* **1974**
[12] Ng K Y and Awang N 2018 *Environmental Monitoring and Assessment* **190**(2)
[13] Mohamad N S, Deni S M and Ul-Saufie A Z 2018 *Pertanika Journal of Science and Technology* **26**(1) 367
[14] Nazif A, Mohammed N I, Malakahmad A, Abualqumboz M S 2018 *Environmental Science and Pollution Research* **25**(1) 283
[15] Fong S Y, Abdullah S and Ismail M 2018 *J of Sustainability Science and Management* **13**(5) 3–17
[16] Afzali A, Rashid M, Afzali M and Younesi V 2017 *J of Cleaner Production* **166** 1216-1225
[17] Mohamad N, Mohddeni S and Ul-Saufie A Z 2017 *J of Environmental Science and Technology* **10(2)** 96-106
[18] Hamid H A, Yahaya A S, Ramli N A, Ul-Saufie A Z and Yasin M N 2016 *MATEC Web of Conferences* **47** 05001
[19] Nazif A, Mohammed N I, Malakahmad A and Abualqumboz M S 2016 *J Water, Air, and Soil Pollution* **227**(4)
[20] Ahmat H, Yahaya A S and Ramli N A 2015 *J of Engineering Science and Technology* **10**(12)

IOP Conf. Series: Earth and Environmental Science **616** (2020) 012008 doi:10.1088/1755-1315/616/1/012008

1560
[21] Sapini M L, Rahim, N Z B A and Noorani M S M 2015 *AIP Conference Proceedings* **1682**
[22] Ahmat H, Yahaya A S and Ramli N A 2015 *J Teknologi* **77**(32) 37
[23] Ul-Saufie A Z, Yahaya A S, Ramli N A, Rosaida N and Hamid H A *2013 J Atmospheric Environment* **77** 621
[24] Hamid H A, Yahaya A S, Ramli N A and Ul-Saufie A Z 2013 *J of Applied Sciences* **13**(2) 294
[25] Noor N M, Tan C Y, Ramli N A, Yahaya A S and Yusof N F F M 2011 *Australian Journal of Basic and Applied Sciences* **5**(12) 2796
[26] Sansuddin N, Ramli N A, Yahaya A S, Yusof N F F M, Ghazali N A and Madhoun W A Al 2011 *Environmental Monitoring and Assessment* **180**(1–4) 573
[27] Md Yusof N F F, Ramli N A, Yahaya A S, Ghazali N A and Al Madhoun W 2010 *J Environmental Monitoring and Assessment* **163**(1-4) 655
[28] Kamarul Zaman N A F, Kanniah K D, and Kaskaoutis D G 2018 IOP *Conference Series: Earth and Environmental Science* **169**(1) 117
[29] Ul-Saufie A Z, Yahaya A S, Ramli N A and Hamid H A 2015 *J of Environmental Science and Technology* **8**(2) 59
[30] Afzali A, Rashid M, Sabariah B and Ramli M 2014 *IOP Conference Series: Earth and Environmental Science* **18**(1)