Hybrid Modelling for PM_{10} Prediction in Malaysia during High Particulate Events

Izzati Amani Mohd Jafri[1,2], Norazian Mohamed Noor[1,2], Ahmad Zia Ul-Saufie[2,3], Nur Alis Addiena A Rahim[1,2], Norazrin Ramli[1,2], Mohd Remy Rozainy Mohd Arif Zainol[4], Gyorgy Deak[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 Technology (CEGeoGTech), Jejawi 02600 Arau, Perlis, Malaysia.*
[3]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi Mara (UiTM), Shah Alam 40450, Malaysia*
[4]*School of Civil Engineering, Universiti Sains Malaysia, Engineering Campus, Nibong Tebal, Penang 14300, Malaysia.*
[5]*National Institute for Research and Development in Environmental Protection INCDPM, Splaiul Independentei 294, 060031 Bucharest, Romania*

Abstract

Air pollution remains as a pressing issue worldwide, with particulate matter (PM_{10}) poses as a significant threat to human health and the environment. The increasing frequency of high particulate events (HPE) particularly in Malaysia has highlighted the necessity for a reliable and accurate predicting tool. This study proposed a hybrid method for predicting PM_{10} concentrations during HPE in peninsular Malaysia specifically in Shah Alam, Nilai, Bukit Rambai and Larkin. Hourly air pollutant concentration (PM_{10}, NO_x, NO_2, SO_2, CO, O_3) and meteorological parameters (relative humidity, temperature and wind speed) during the HPE events in 1997, 2005, 2013 and 2015 were used. The hybrid approach applied the combination of Quantile Regression (QR) model and three feature selection methods which is Support Vector Machine (SVM), Correlation (C) and Deviation (D). The optimal QR percentile and selected input parameter from each feature selection methods were applied to develop the prediction models. Then, the model performance was evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Index of Agreement (d_2). Findings show that SVM-QR and C-QR model consistently provided the best accuracy at all locations with the range of MAE value (5.66-11.89) for the next-day PM_{10} prediction. Validation using the unseen dataset from 2019 also confirmed the strong alignment between observed and predicted value. In conclusion, the proposed hybrid models demonstrated a reliable performance and offering a practical tool for predicting PM_{10} concentration during HPE and supporting air quality management in Malaysia.

Keywords: Air Quality; High Particulate Event; Hybrid Modeling; Quantile Regression; Feature Selection