

# Evaluation of Machine Learning Models for Estimating PM<sub>2.5</sub> Concentrations across Malaysia

Nurul Amalin Fatihah Kamarul Zaman <sup>1</sup>, Kasturi Devi Kanniah <sup>1,2,\*</sup>, Dimitris G. Kaskaoutis <sup>3,4,\*</sup>  
and Mohd Talib Latif <sup>5</sup>

**Citation:** Zaman, N.A.F.K.; Kanniah, K.D.; Kaskaoutis, D.G.; Latif, M.T. Evaluation of Machine Learning Models for Estimating PM<sub>2.5</sub> Concentrations across Malaysia. *Appl. Sci.* **2021**, *11*, 7326. <https://doi.org/10.3390/app11167326>

<sup>1</sup> Tropical Map Research Group, Faculty of Built Environment & Surveying, Universiti Teknologi Malaysia, Skudai 81310, Johor, Malaysia; nafatihah4@graduate.utm.my

<sup>2</sup> Centre for Environmental Sustainability and Water Security (IPASA), Research Institute for Sustainable Environment, Universiti Teknologi Malaysia, Skudai 81310 UTM, Johor, Malaysia

<sup>3</sup> Institute for Environmental Research and Sustainable Development, National Observatory of Athens, 15236 Athens, Greece

<sup>4</sup> Environmental Chemical Processes Laboratory, Department of Chemistry, University of Crete, 71003 Crete, Greece

<sup>5</sup> Department of Earth Sciences and Environment, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia; talib@ukm.edu.my

\* Correspondence: kasturi@utm.my (K.D.K.); dkask@noa.gr (D.G.K.)

Academic Editor: Yves Rybarczyk

Received: 25 June 2021

Accepted: 5 August 2021

Published: 9 August 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

## Supplementary Materials

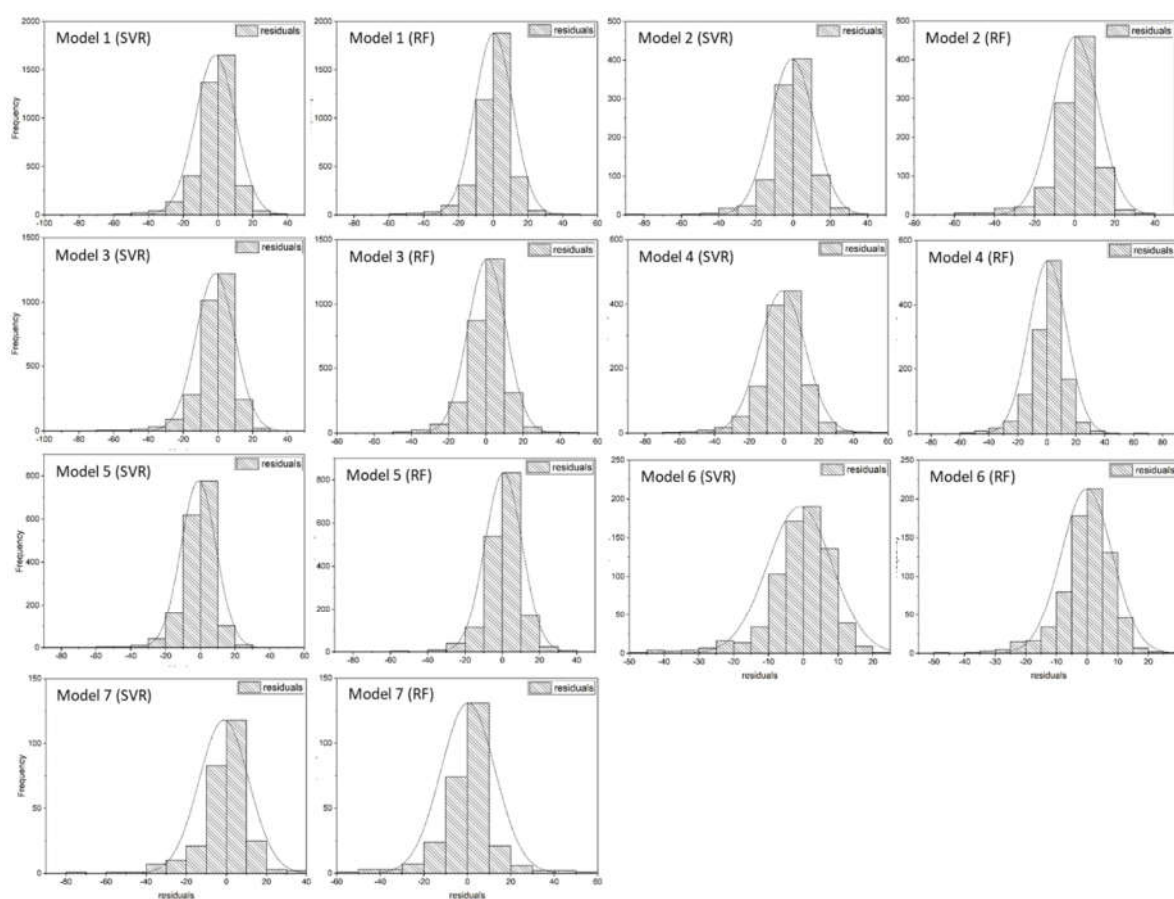
**Table S1.** T Parameters/functions used for the SVR model.

| Function  | Default values     |
|---|--------------------|
| Radial basis function (RBF) kernel with gamma parameter | 1/(data dimension) |
| cost  | 1                  |
| Epsilon   | 0.1                |

**Table S2.** Coefficient of determination ( $R^2$ ), RMSE, MBE and NSE values using SVR and RF models for PM<sub>2.5</sub> estimations in Malaysia. The statistical indicators are presented as averaged values for each model and Number of samples.

| Model | Technique | Calibration dataset |       |        |       |                  | Validation dataset |       |        |       |                  | Importance ranking   |
|-------|-----------|---------------------|-------|--------|-------|------------------|--------------------|-------|--------|-------|------------------|--|
|       |           | $R^2$               | RMSE  | MBE    | NSE   | No of sample (N) | $R^2$              | RMSE  | MBE    | NSE   | No of sample (N) |  |
| 1     | SVR       | 0.69                | 10.62 | -1.392 | 0.679 | 9365             | 0.66               | 12.11 | -1.619 | 0.639 | 4011             | CO, AOD, O <sub>3</sub> , NO <sub>2</sub> , SO <sub>2</sub> , RH, WD, TEMP, WS |
|       | RF        | 0.66                | 11.28 | 0.066  | 0.939 | 9363             | 0.62               | 11.40 | 0.097  | 0.647 | 4013             |  |
| 2     | SVR       | 0.79                | 10.23 | -1.191 | 0.777 | 2352             | 0.67               | 12.47 | -0.709 | 0.661 | 1005             | CO, AOD, O <sub>3</sub> , NO <sub>2</sub> , RH, SO <sub>2</sub> , WD, WS, TEMP |
|       | RF        | 0.67                | 12.21 | 0.063  | 0.944 | 2349             | 0.76               | 11.47 | 0.346  | 0.735 | 1008             |  |
| 3     | SVR       | 0.69                | 10.67 | -1.416 | 0.668 | 6859             | 0.61               | 11.53 | -1.529 | 0.591 | 2939             | CO, O <sub>3</sub> , AOD, NO <sub>2</sub> , SO <sub>2</sub> , TEMP, RH, WD, WS |
|       | RF        | 0.64                | 11.19 | 0.071  | 0.936 | 6858             | 0.64               | 10.76 | 0.479  | 0.629 | 2940             |  |
| 4     | SVR       | 0.81                | 11.18 | -1.187 | 0.794 | 2960             | 0.67               | 14.00 | -1.226 | 0.671 | 1263             | CO, AOD, O <sub>3</sub> , NO <sub>2</sub> , SO <sub>2</sub> , TEMP, RH, WD, WS |
|       | RF        | 0.72                | 12.77 | 0.085  | 0.952 | 2956             | 0.71               | 13.61 | 0.524  | 0.709 | 1267             |  |
| 5     | SVR       | 0.62                | 9.96  | -1.437 | 0.599 | 4060             | 0.50               | 10.90 | -1.168 | 0.492 | 1737             | CO, O <sub>3</sub> , AOD, NO <sub>2</sub> , TEMP, SO <sub>2</sub> , RH, WD, WS |
|       | RF        | 0.55                | 10.58 | 0.047  | 0.917 | 4057             | 0.56               | 9.99  | 0.684  | 0.558 | 1740             |  |
| 6     | SVR       | 0.58                | 8.62  | -0.994 | 0.555 | 1709             | 0.46               | 9.38  | -1.008 | 0.450 | 732              | CO, NO <sub>2</sub> , O <sub>3</sub> , WS, SO <sub>2</sub> , TEMP, AOD, RH, WD |
|       | RF        | 0.45                | 9.75  | 0.066  | 0.899 | 1708             | 0.53               | 8.36  | -0.091 | 0.524 | 733              |  |
| 7     | SVR       | 0.74                | 11.17 | -2.323 | 0.687 | 643              | 0.59               | 12.46 | -1.243 | 0.568 | 272              | CO, AOD, O <sub>3</sub> , NO <sub>2</sub> , WS, SO <sub>2</sub> , TEMP, WD, RH |
|       | RF        | 0.66                | 11.54 | -0.108 | 0.945 | 640              | 0.59               | 12.41 | 0.331  | 0.578 | 275              |  |

Notes: Model 1: Overall model; Model 2: Spatial model (urban/industrial); Model 3: Spatial model (suburban); Model 4: Temporal model (dry season); Model 5: Temporal model (wet season); Model 6: Temporal model (inter-monsoon April-May); Model 7: Temporal model (inter-monsoon October).



**Figure S1.** Residual analysis (residuals = predicted PM<sub>2.5</sub> - measured PM<sub>2.5</sub>) from SVR and RF for the developed Models 1–7. The fitted curve represents the normal distribution.