SIIC073

PREDICTION OF TOTAL MAXIMUM DAILY LOADS (TMDLs) OF POLLUTANTS IN RIVER BY USING ARTIFICIAL NEURAL NETWORK (ANN)

Khairunnisa Khairudin[1], Mohamed Syazwan Osman[1], and Syahrul Fithry Senin[2]
[1]*Faculty of Chemical Engineering, Universiti Teknologi MARA Pulau Pinang, 13500 Permatang Pauh, Pulau Pinang Malaysia*
[2] *Faculty of Civil Engineering, Universiti Teknologi MARA Pulau Pinang, 13500 Permatang Pauh, Pulau Pinang Malaysia*

**Corresponding author: khairunnisakhairudin96@gmail.com*

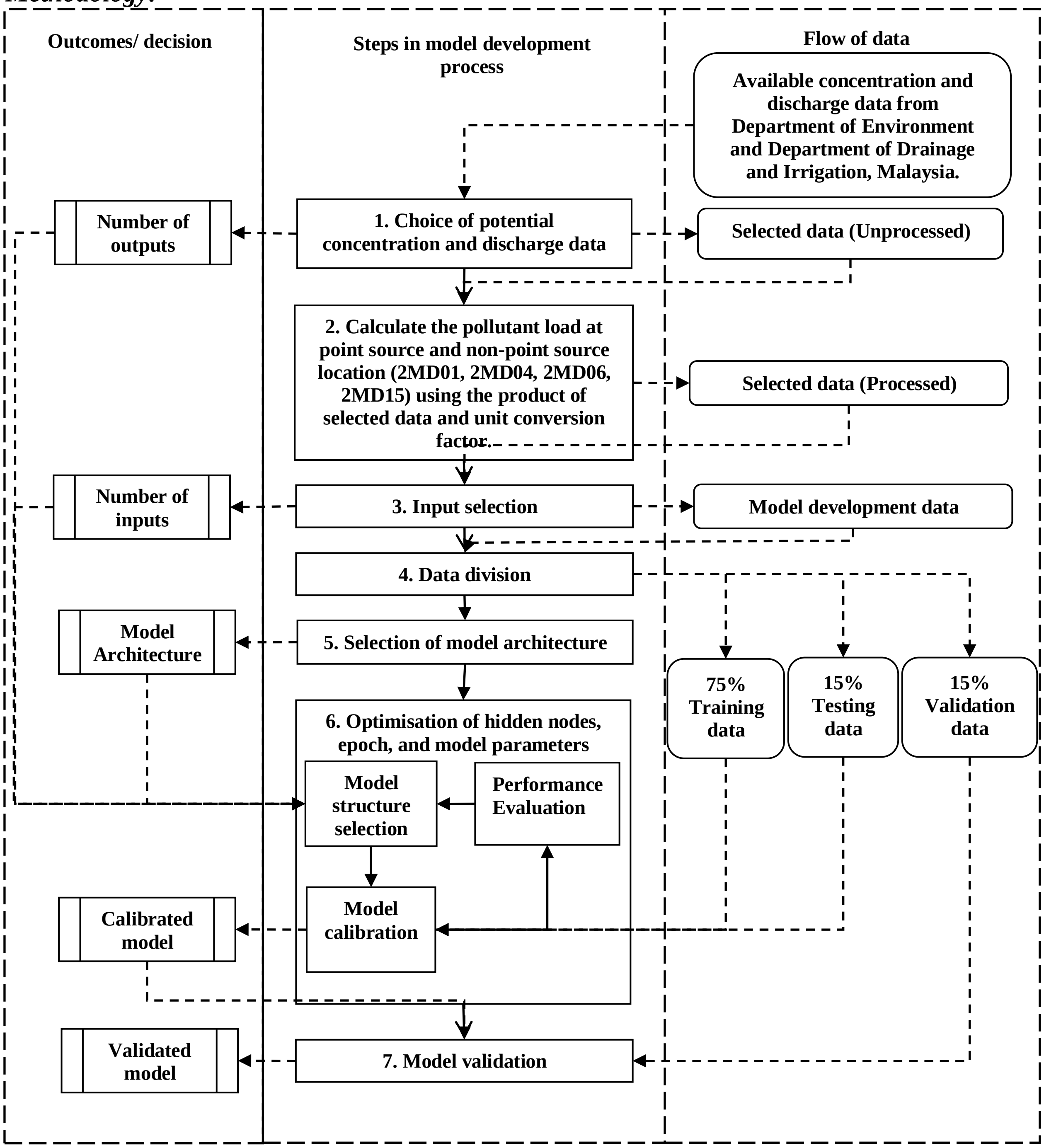
Abstract:
Total Maximum Daily Load (TMDL) studies are crucial in determining a pollutant reduction target and allocates load reductions necessary to the source(s) of the pollutant. Existing modelling approaches to simulate TMDL allocations of point source and non-point source pollutants typically consist of linking watershed model, receiving water transport model, and receiving water quality model. Such deterministic model requires extensive data of the underlying process compared to artificial neural network (ANN) that simulates data based on data-driven method. In this study, biochemical oxygen demand (BOD), chemical oxygen demand (COD), suspended solids (SS), and ammoniacal nitrogen (NH_3-N) loads for Muda River is predicted using ANN. The model is developed based on historical monthly concentration data and discharge data from 2013 to 2018 provided by Department of Environment (DOE), Malaysia. These parameters were introduced as inputs, whereas TMDL as outputs of the three-layer feed-forward back-propagation ANN. The learning algorithm used is Bayesian Regularization with tansig transfer function at the hidden layer and purelin transfer function at the output layer. Here, the number of neurons tested to obtain the optimum number of hidden layer nodes is 5, 7, 9, 11, and 13, which run at different epochs: 1000, 2000, and 3000. Model performance was evaluated using mean absolute percent error (MAPE), coefficient of determination (R^2), root mean square error (RMSE), and model efficiency (E). The best model for TMDL of BOD is 6:13:1 at epoch 2000 with 0.0004% (MAPE), 1.0 (R^2), 0.0005 (RMSE), and 1.0 (E). Meanwhile, the best model for TMDL of COD is 6:5:1 at epoch 3000 with 0.00004% (MAPE), 1.0 (R^2), 0.0004 (RMSE), and 1.0 (E). Furthermore, the best model for TMDL of SS is 6:5:1 at epoch 3000 with 0.0038% (MAPE), 0.99 (R^2), 0.1 (RMSE) and 1.0 (E). Finally, the best model for TMDL of NH_3-N is 6:5:1 at epoch number 3000 with 0.0001% (MAPE), 1.0 (R^2), 9.47×10^{-6} (RMSE) and 1.0 (E). It can be concluded that ANN is an excellent modelling approach to substitute deterministic models for TMDL prediction.

***Keywords*:**
TMDL study; Artificial neural network (ANN); Water quality parameters; *Bayesian regularization*.

Objectives:

- To determine the total maximum daily loads of biochemical oxygen demand, chemical oxygen demand, ammoniacal nitrogen, and suspended solids.
- To compare and validate the simulated data with field data obtained.

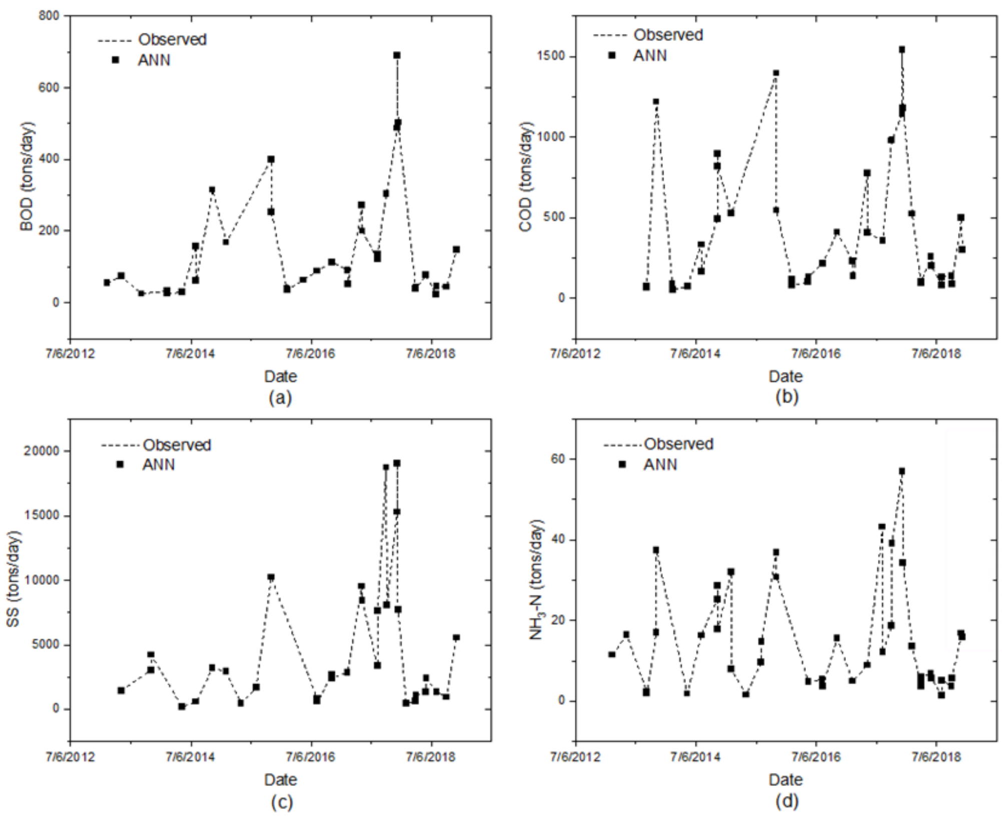
Methodology:



H. R. Maier, A. Jain, G. C. Dandy and K.P.Sudheer, "Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions," Environmental Modelling and Software, vol. 25, pp. 891-909, 2010.

Results:

Time-Series Plot of Predicted versus Observed TMDL. (a) BOD Model (b) COD Model (c) SS Model (d) NH_3-N Model



Conclusion:

As a conclusion, the total maximum daily loads can be successfully predicted using the neural network. The TMDL model for BOD requires lower iteration numbers and high neuron numbers in the hidden layer compared to the TMDL model for COD, NH_3-N, and SS requiring high iteration numbers and low hidden nodes. This is due to the complexity of the data input into the neural networks. The hidden nodes and epoch number play a significant role in optimizing the neural network model as insufficient hidden layer neuron number and epoch number may lead to underfitting, while excessive hidden nodes number and epoch number may lead to overfitting of the model. It can also be concluded that the feed-forward back-propagation neural network with Bayesian Regularization training function is capable of recognizing the pattern of pollutant load at point source and non-point source and of providing accurate TMDL predictions for BOD, COD, SS, and NH_3-N. It is proven in the time-series plot of the predicted versus observed TMDL, as it depicts a closely followed pattern of variation by the observed and predicted TMDL values. Finally, the developed neural network TMDL model shows an accurate prediction as it has a lower error between observed and predicted TMDL. Artificial neural network is therefore an excellent alternative to the current process-based model for TMDL prediction. The model obtained may be used to predict TMDL in the river basin. For further research, this model can be compared with a different model to justify the reliability of the FFBP-NN model.