Comparison of Artificial Neural Network and Support Vector Machine for Long-Term Runoff Simulation

Z Hassan, S Z Rosdi, A N Kamarudzaman, M Abdul Rahim and Z M Ghazaly

Civil Engineering Program, School of Environmental Engineering, Universiti Malaysia Perlis (UniMAP), 02600 Arau, Perlis, Malaysia.

Email: zulkarnainh@unimap.edu.my

Abstract. Simulation of runoff from a river catchment is a very difficult task and it involves a lot of data which need to be considered. However, the modelling is very essential to forecast the patterns of future runoff by observing and analysing previous patterns of runoff, based on the rainfall. This study presents the evaluation of rainfall-runoff modelling for the long-term runoff series using Artificial Neural Network (ANN) and Support Vector Machine (SVM). Both models are trained and validated using the data series of current and nine (9) antecedent rainfall. During the validation, the SVM model is better in the performance as compared the ANN model, with the R and RMSE values are 0.529-0.711 and 14.27-52.55 mm, respectively. However, the SVM model is underestimated for the peak discharge. Both models have the ability to derive the relationship between the inputs and outputs of the rainfall-runoff process.

1. Introduction

One of the fundamental fields of hydrology is the rainfall-runoff relationship, which plays an important role in water balance [1]. This non-linear relationship for the runoff prediction depends on many factors such as rainfall intensity and duration, and physical characteristics of the upstream catchment. It involves a lot of data and factors which need to be analysed. The conventional of rainfall-runoff modelling has always been a very difficult task. However, the modelling is very essential in order to forecast the patterns of future runoff by observing and analysing previous patterns of runoff, based on the rainfall parameters [2]. In this study, matrix-based models in which alternative to the conventional are been applied, which are Artificial Neural Network (ANN) and Support Vector Machine (SVM).

The ANN model has been satisfactorily applied to the prediction of nonlinear hydrologic processes such as rainfall-runoff, streamflow, precipitation, and water quality modelling [3]. The model is always been select as robust matrix-based models for simulating the runoff (or discharge) and been used to compare with other models [4]. Recently, there have been very significant developments in investigated the potential of SVMs in modelling runoff based on rainfall inputs. A few explorations on the SVM model over the entire area of Malaysia have been recorded such as [2] and [5]. However, comparison and evaluation between the ANN and SVM model models have not been well explored. Therefore, this evaluation can be regarded as a research interest in this study in relation to the studied region. Hence, the aim of this study to investigate the applicability of ANN and SVM in the prediction of rainfall-runoff by using the minimum input information such as daily rainfall and antecedent rainfall. The performance of both models will be evaluated.
2. Materials and methods

2.1. Study area

In this study, Muda River is selected as a study area as shown in Figure 1. This river is located in Kedah, Malaysia and the catchment area of the river is 4 219 km². The length of the Muda river is about 178 km which is the longest river in Kedah. All hydrological data (Table 1) such as rainfall and discharge from 1980-2016 are provided by the Department of Irrigation and Drainage (DID), Malaysia. The data are homogenized, and the missing data have been treated. In Catchment D, PRE5808001 has been used as the rainfall data for the catchment since there is no rainfall gauge in the catchment.

![Figure 1. Location of rainfall and discharge gauges.](image-url)

| Sub-Catchments | Station No. | Type of Data | Station Name | Lat. (° ' '' N) | Long. (° ' '' E) |
|----------------|-------------|--------------|--------------|-----------------|-----------------|
| A              | PRE6108001  | Rainfall     | Komplek Rumah Muda, Kedah | 06 06 20 | 100 50 50 |
|                | DIS5806414  | Discharge    | Sg. Muda at Jeniang, Kedah | 05 49 10 | 100 37 55 |
| B              | PRE5507076  | Rainfall     | Bt.27 Jln. Baling, Kedah | 05 35 00 | 100 44 10 |
|                | DIS5606410  | Discharge    | Sg. Muda at Jam. Syed Omar, Kedah | 05 36 35 | 100 37 35 |
| C              | PRE5808001  | Rainfall     | Bt.61 Jln. Baling, Kedah | 05 52 50 | 100 53 40 |
|                | DIS5608418  | Discharge    | Sg. Ketil at Kuala Pegang, Kedah | 05 38 20 | 100 48 45 |
| D              | DIS5505412  | Discharge    | Sg. Muda at Ldg. Victoria, Penang | 05 31 55 | 100 34 20 |
2.2. Description of rainfall-runoff models

2.2.1. Multilayer Perceptron (MLP) of Artificial Neural Network. Multilayer perceptron network (MLP) consists of input, hidden, and output layers with their nodes and activation functions. For a network training method, the back-propagation algorithm (BPA) introduced by [6] and it can effectively train the network for nonlinear problems, which has stimulated a torrent of research in neural networks. The activation function consists of a log-sigmoid function in the hidden layer and a linear function in the output layer. It has been reported that ANNs with this configuration are the most commonly used form, as they have improved extrapolation ability. The detailed algorithm of MLP can be referred to at [7].

2.2.2. Support Vector Machine (SVM). SVM has been developed based on the statistical learning theory aiming at minimizing the generalized model error rather than just minimizing the training error. This minimizing which consequently increases SVM generalization ability [8]. SVM has been developed as a classification tool and it was applied successfully in a wide range of classification and clustering applications in.

In this study, Support Vector Regression (SVR) of SVM is used. The function of SVR is transforming the input data by linear or nonlinear mapping into a high-dimensional feature space. The architecture of SVR can be detailed as shown in Figure 2, where $K(x, x_i)$ is the output of the $i$-th hidden node for input vector $x$, it is a mapping of the input $x$ and the support vector $x_i$ by selecting the kernel function [9].

![Figure 2. SVR architecture][2]

2.3. Training and Validation of the rainfall-runoff models

To develop these two models, the data are divided into two periods, which are named as training (1980-2008) and validation (2009-2016). The input data used for both models consist of current ($\{P(t)\}$) and antecedent ($\{P(t-1), P(t-2), \ldots, P(t-n)\}$) daily rainfall, which only nine of the antecedent rainfall are used in the model development. The antecedent rainfall is proposed to use to reduce the problem of pattern recognition in the ANN and SVM models.

The performance of both models is checked based on the criterion of visual comparison between observed and simulated data, that involves obtaining the best match between the simulated and observed hydrograph. The correlation coefficient ($R$) (Eq. 1) and root mean square error (RMSE) (Eq. 2) are also used to measure the model performance, where $N$ is the number of samples, $x$ and $y$ are the
observed and predicted values for \(i=1, \ldots, n\) and \(x\) and \(y\) are the mean values of the observed and predicted data set, respectively. A better agreement between observed and predicted values are expressed by the \(R\) and \(RMSE\) values as close to 1 and 0, respectively.

\[
R = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}}
\]

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N}(x_i - y_i)^2}
\]

3. Results and discussion

The ANN and SVM models as rainfall-runoff models are carried out on the daily rainfall series. During the model development, rainfall data are separated into 29-year (training) and 8-year (validation). The training and validation sets of rainfall data are separated into 18-year and 2-year, respectively. Combination of current and nine (9) antecedence rainfall series are used as the input for both models. The overall results of the model development of the models are shown in Table 2 and Figures 3-6.

| Table 2. Performance (best in bold) during training and validation periods for the ANN and SVM models |
| --- |
| **Model Development** | **Discharge Station** | **ANN** | **SVM** |
|  |  | **R** | **RMSE (mm)** | **R** | **RMSE (mm)** |
| Training | DIS5806414 | 0.737 | 24.59 | 0.518 | 31.69 |
|  | DIS5606410 | 0.769 | 58.7 | 0.632 | 74.86 |
|  | DIS5608418 | 0.647 | 24.26 | 0.412 | 30.77 |
|  | DIS5505412 | 0.734 | 71.4 | 0.609 | 86.64 |
| Validation | DIS5806414 | 0.524 | 30.89 | 0.529 | 29.02 |
|  | DIS5606410 | 0.703 | 52.81 | 0.711 | 47.79 |
|  | DIS5608418 | 0.421 | 21.94 | 0.609 | 14.27 |
|  | DIS5505412 | 0.685 | 65.98 | 0.682 | 52.55 |

As shown in Table 2, the ANN model shows the best performance during the training period for all discharge stations as compared to the SVM models. The range of ANN performances for the ANN model during the training are 0.647-0.769 and 24.26mm-71.4mm for the \(R\) and \(RMSE\) values, respectively as compared to the SVM model with the 0.412-0.632 and 30.77-86.64mm for the \(R\) and \(RMSE\) values, respectively. The SVM model is able to give a small difference in simulating the runoff, it shows a small value of \(R\) and \(RMSE\), which indicated the good performance as compared the ANN model during the validation period, with the \(R\) values of 0.529-0.711 and the \(RMSE\) values of 14.27-52.55 mm. Although the ANN can perform better in calibration periods, SVM shows drastically well performed during validation periods. The parameter of ANN needs to be adjusted and the performance of SVM model depends on the choice of the kernel function and the model parameters in order to gain the optimum and efficient result [10].

The performance of the models is also shown by visual comparison between observed and simulated discharge during the validation period as detailed in Figures 3-6. As shown in those figures, it seems that the SVM model is able to capture the observed mean and low discharge as compared to the ANN model. This pattern of the result similar with [11], in which the SVM model has a higher non-linear mapping capability and allowing it to detect runoff data patterns (in this study, observed mean) more easily than the ANN model. However, the SVM model is unable to detect well the for the most of peak discharge. It is be noticed that the ANN model can capture well the peak discharge. From those results, even though the SVM model is able to simulate well in the statistic comparison (Table 2), the simulated SVM is unable to capture well the observed peak discharge for all discharge stations as shown in Figures 3-6.
Figure 3. Performance of the (a) ANN and (b) SVM model during the validation period at DIS5806414 (Sub-Catchment A)

Figure 4. Performance of the (a) ANN and (b) SVM model during the validation period at DIS5606410 (Sub-Catchment B)

Figure 5. Performance of the (a) ANN and (b) SVM model during the validation period at DIS5608418 (Sub-Catchment C)
4. Conclusion
The performance of ANN and SVM are evaluated for discharge simulation with a long daily rainfall series. The data from Muda river, Kedah are used as a case study and the combination of input of present and nine-antecedent rainfall are used in order to evaluate the performance of the ANN and SVM. For the overall results, the ANN is performed well in the calibration period, but the performance decreased in validation periods. While SVM gives the better performance in validation rather than in calibrations periods for each station where the R values getting higher and RMSE values getting lower. However, the SVM model is unable to capture the peak discharge and this result is vice versa for the ANN model. These performances reveal that both models are giving a remarkable performance if the models customize with a good configuration. Further studies must be conducted for generalizations of the results.

Acknowledgment
This work had been financially supported by the Ministry of Education Malaysia, under the Fundamental Research Grant Scheme (FRGS) (Ref. No.: FRGS/1/2019/TK01/UNIMAP/02/4). The authors would like to thank the Department of Irrigation and Drainage Malaysia (DID) for providing data and technical support.

References
[1] Li H, Zhang Y and Zhou X 2015 Advances in Meteorology 2015 720967
[2] Seyam M, Othman F and El-Shafie A 2017 MATEC Web of Conferences 111 01007
[3] ASCE 2000 Artificial Neural Networks in hydrology I: Preliminary concepts Journal of Hydrologic Engineering 5 2 pp 115-123
[4] Kisi O, Nia A M, Goshen M G, Tajabadi M R J and Ahmadi A 2012 Water resources management 26 2 pp 457-474.
[5] Zaini N, Malek M A, Yusoff M, Mardi N H and Norhisham S 2018 IOP Conference Series: Earth and Environmental Science 140 1 012035.
[6] Rumelhart D E, Hinton G E and Williams R J 1985 California Univ San Diego La Jolla Inst for Cognitive Science.
[7] Hassan Z, Shamsudin S, Harun S, Malek M A and Hamidon N 2015 Environmental Earth Sciences 74 1 pp 463–477
[8] Behzad M, Asghari K and Coppola Jr E A 2009 Journal of Computing in Civil Engineering 24 5 pp 408-413.
[9] Chen S T and Yu P S 2007 Journal of Hydrology 347 1-2 pp 67-78.
[10] Botsis D, Latinopoulos P and Diamantaras K 2011 In 12th International Conference on
Environmental Science and Technology (CEST2011) pp 8-10.

[11] Hu C H, Wang J J and Wu Z N 2011 In 2011 International Symposium on Water Resource and Environmental Protection 2 pp 1099-1103