Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 154 (2016) 1103 – 1109

Procedia Engineering

www.elsevier.com/locate/procedia

12th International Conference on Hydroinformatics, HIC 2016

Rainfall-runoff modeling using Dynamic Evolving Neural Fuzzy Inference System with online learning

Chang Tak Kwin[a], Amin Talei[a,]*, Sina Alaghmand[a], Lloyd H. C. Chua[b]

[a]*School of Civil Engineering, Monash University Malaysia, Jalan Lagoon Selatan, Bandar Sunway, 47500 Selangor Darul Ehsan, Malaysia*
[b]*School of Engineering, Deakin University, Locked Bag 20000, Geelong, VIC 3220, Melbourne, Australia*

Abstract

Neuro-Fuzzy Systems (NFS) are computational intelligence tools that have recently been employed in hydrological modeling. In many of the common NFS the learning algorithms used are based on batch learning where all the parameters of the fuzzy system are optimized off-line. Although these models have frequently been used, there is a criticism on such learning process as the number of rules are needed to be predefined by the user. This will reduce the flexibility of the NFS architecture while dealing with different data with different level of complexity. On the other hand, online or local learning evolves through local adjustments in the model as new data is introduced in sequence. In this study, dynamic evolving neural fuzzy inference system (DENFIS) is used in which an evolving, online clustering algorithm called the Evolving Clustering Method (ECM) is implemented. ECM is an online, maximum distance-based clustering method which is able to estimate the number of clusters in a data set and find their current centers in the input space through its fast, one-pass algorithm. The 10-minutes rainfall-runoff time series from a small (23.22 km^2) tropical catchment named Sungai Kayu Ara in Selangor, Malaysia, was used in this study. Out of the 40 major events, 12 were used for training and 28 for testing. Results obtained by DENFIS were then compared with the ones obtained by physically-based rainfall-runoff model HEC-HMS and a regression model ARX. It was concluded that DENFIS results were comparable to HEC-HMS and superior to ARX model. This indicates a strong potential for DENFIS to be used in rainfall-runoff modeling.

Peer-review under responsibility of the organizing committee of HIC 2016

Keywords: Rainfall-runoff modeling; Neuro-fuzzy systems; DENFIS; ARX; HEC-HMS

* Corresponding author. Tel: +60355145648; fax: +60355146207.
E-mail address: amin.talei@monash.edu

Peer-review under responsibility of the organizing committee of HIC 2016
doi:10.1016/j.proeng.2016.07.518

1. Introduction

Rainfall-runoff (R-R) modeling is a popular field of study in hydrology research which aims to understand and capture the rainfall-runoff process. The models commonly used for R-R modeling can be categorized into two groups: (1) physically-based models such as Storm Water Management Model (SWMM) and Hydrologic Engineering Center – Hydrologic Modeling System (HEC-HMS); (2) system theoretic models such as Artificial Neural Networks (ANN) and Neuro-Fuzzy Systems (NFS). Neuro-Fuzzy Systems are hybridizations of fuzzy theory sets with neural networks which maps the relationship between the input and output data through the fuzzy inference system (FIS). There are generally two classes of FIS, linguistic or precise models. Of the two, precise models or Takagi-Sugeno-Kang (TSK) models are more readily used in R-R modeling. Precise models are models where only the rule antecedent is defined by a fuzzy set while the consequent can be represented by crisp values. A typical fuzzy inference system generally includes four main components: (1) fuzzification interface for transforming crisp values into linguistic fuzzy sets; (2) an interface engine for rule inferencing; (3) fuzzy rule base containing fuzzy IF-THEN rules; and (4) defuzzification interface which transforms the output into crisp values.

NFS learning is usually classified as either offline or online learning. Online learning allows models to dynamically reiterate parameters based on new training data point that are presented to the model while offline learning or batch learning optimizes parameters based on a static dataset. Adaptive Network-based Fuzzy Inference System (ANFIS) [1] is a popular model used in R-R modeling; however, ANFIS employs offline learning which suffers from increased computational time and requires retraining to capture recent changes of the system. Moreover, the number of rules in ANFIS is predefined by the user. Dynamic Evolving Neural-Fuzzy Systems (DENFIS) [2] is a TSK type model which employs online learning capabilities that allows dynamic reiteration of the model parameters. This allows DENFIS to dynamically evolve its rule base to capture changes within the system through continuous updating of the model. To date, there are only few studies on NFS with online learning in R-R modeling. This study aims to investigate the application NFS with online learning for event-based rainfall-runoff modeling for a rural tropical catchment of size 23.22km^2. The results obtained would be compared against an autoregressive model with exogenous inputs (ARX) and a physically-based model, Hydrologic Engineering Center – Hydrologic Modeling System (HEC-HMS).

2. Dynamic Evolving Neural-Fuzzy Inference System (DENFIS)

DENFIS is an adaptive online model capable of incremental learning that uses the Takagi-Sugeno type fuzzy inference system [2]. DENFIS employs an evolving, online clustering method known as Evolving Clustering Method (ECM). ECM is a fast, one-pass algorithm proposed to implement as a scatter partitioning of input space during the creating of fuzzy inference rules. It dynamically estimates the number of clusters and its centers within a set of data. ECM is a distance-based connectionist clustering method where the number of clusters is estimated using the maximum distance between a point and a cluster center. The parameter, D_{thr}, is a threshold value for maximum calculated distance, hence it allows the numbers of clusters to be controlled. The general Euclidean distance between vectors x and y is denoted as below:

$$\|x - y\| = \left(\sum_{i=1}^{q} |x_i - y_i|^2 \right)^{\frac{1}{2}} \Big/ q^{\frac{1}{2}} \tag{1}$$

where $x, y \in R^q$ and q is the number of input data points. As more data is added into the system, either new clusters may form or existing clusters may get updated in terms of cluster center and cluster size. The cluster update process ends when its cluster radius reaches a threshold value of D_{thr}. Fig. 1 shows the clustering process in 2-D space.

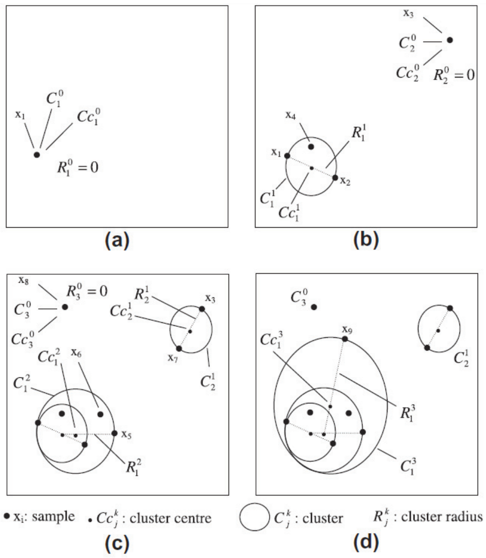


Fig. 1. Schematic of ECM clustering procedure using the samples x_1 to x_9 in 2-D space [2].

The ECM algorithm is summarized as follows:

Step 0: The first cluster C_1^0 is created by the first training tuple as the first cluster center $C_{C1}^{\ 0}$ with cluster radius $R_1 = 0$ (Fig. 1a).

Step 1: If all examples of the data stream have been presented, the algorithm is complete. Else the distance between the current example and all n already created cluster centers $C_{Cj}, D_{Cj} = \|x_i - C_{Cj}\|$ for $j = 1,2,\ldots,n$ are calculated where j is the cluster index.

Step 2: The distance D_{ij} is calculated and compared against all existing cluster radius R_j. If at least one radii satisfies the condition $D_{ij} \le R_j$, then x_i belongs to a cluster C_m with the minimum distance $D_{im} = \|x_i - C_{Cm}\| = \min\left(\|x_i - C_{Cj}\|\right)$ for $D_{ij} \le R_j\,(j = 1,2,\ldots,n)$. In this case, no new cluster is either created or updated (the cases of x_4 and x_6 in Fig. 1Fig. 1b and 1c); the algorithm returns to Step 1, else the algorithm continues to the next step.

Step 3: For all n existing cluster centers, find cluster C_a through calculating the values $S_{ij} = D_{ij} + R_j$ for j = 1,2,…,n, and then choosing the cluster center C_{Ca} with a minimum value S_{ia}.

Step 4: If S_{ia} is greater than $2 \times D_{thr}$, the example x_i does not belong to any existing clusters. Hence a new cluster is created similarly to step 0 (the cases of x_3 and x_8 in Fig. 1), the algorithm then returns to step 1.

Step 5: The cluster C_a is updated if S_{ia} is not greater than $2 \times D_{thr}$ by moving its center, C_{Ca}, and increasing its radius, R_a. The updated radius is then set to be equal to $S_{ia}/2$ and the new center is located at the point on the line connecting x_i and C_{Ca}, and the distance from the new center to the point x_i is equal to the updated radius (the cases of x_2, x_5, x_7 and x_9 in Fig. 1). The algorithm proceeds to step 1.

3. Methodology

3.1. Study site and data used

In this study, rainfall and runoff data was collected from the Sungai Kayu Ara river basin. Sungai Kayu Ara river basin covers an area of 23.22 km² as shown as Fig. 2. The main river of this river basin originates from the reserved highland area of Penchala and Segambut. From climatologic aspect, Sungai Kayu Ara river basin lies in equatorial zone. It subjects to northeast monsoon (December to March) and southwest monsoon (June to September). The inter-monsoon seasons normally starts from April to May and from October to November [3]. Annual mean rainfall in this river basin is more than 2000 mm which has been proved by Desa, et al. [4] according to their investigation on Sungai Kerayong river basin in the southeast of Kuala Lumpur about 30 km away. Average daily temperatures range from 25° Celsius to 33° Celsius. The mean monthly relative humidity falls within 70 percent to 90 percent depending upon the location of area and degree of rainfall season. Most of the area in the Sungai Kayu Ara river basin is flattened for development. The annual average evaporation rate for the Sungai Kayu Ara river basin estimated at 4 to 5mm per day. This catchment was chosen since it is a representative of small urbanized catchments in tropical areas. Rainfall-runoff modeling for such catchments should be normally event-based. Moreover, this catchment has got many rainfall stations which makes its input selection and model development quite challenging.

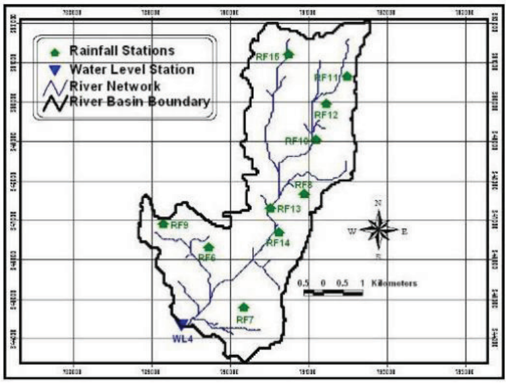


Fig. 2. Schematic map of Sungai Kayu Ara catchment.

3.2. Physically-based model used

Hydrologic Engineering Center – Hydrologic Modeling System (HEC-HMS) is a lumped conceptual model in hydrological applications. It attempts to simulate the physical processes within the rainfall-runoff response of a river basin system to a precipitation input through conceptualizing the entire river basin as a system that is interconnected by hydrologic and hydraulic components like river basins, streams and reservoirs. HEC-HMS is designed to be light in computational complexity but flexible for a wide range of geographic areas with different environment and climates. It has been widely used in many studies involving water resources [5-7]. The model includes many of the processes involved in water circulation in the basin, such as, precipitation, evaporation or infiltration [8]. However, such as every simulation system it has limitations due to the aspects of simplified model formulation and simplified flow representation. The simplification aspects in the model allows quick complete simulations while producing accurate and precise results.

HEC-HMS requires pre-processing through HEC-GeoHMS (Geospatial Hydrologic Modeling). HEC-GeoHMS is an extension of ArcGIS which is specifically designed for surface delineation and producing the required geospatial data for HEC-HMS hydrologic modeling. A surface Digital Elevation Model (DEM) was used to extract drainage paths and watershed boundaries to represent the hydrologic structure used for simulating the watershed response to precipitation. Results produced by HEC-GeoHMS is then extracted that exported into HEC-HMS for watershed hydrologic modeling.

3.3. Input data selection and model development

A total of 40 rainfall-runoff events with hourly data were extracted from available rainfall and runoff time series of Sungai Kayu Ara catchment from which 12 events were used as training events while the remaining 28 were used as testing events. Care was taken in selecting training and testing events so both data set contained low, medium, and high events. Rainfall antecedents from all 10 rainfall stations and runoff antecedents were provided for input selection model. Rainfall antecedents were considered up to 10 lagged time steps as initial results showed that lags beyond this are not contributing much in R-R process.

One main difficulty in the application of data driven models are the choice of inputs. Dinpashoh, et al. [9] suggested several methods for inputs selection analysis in data-driven models. Input selection analysis is required for identifying proper rainfall and/or runoff antecedents to train the model. Correlation analysis is one of the most commonly-used methods to identify proper inputs in data-driven models including NFS models. Moreover, mutual information analysis has been used for input selection in several similar studies and is found to be a good choice in identifying appropriate inputs [10, 11]. Talei and Chua [12] proposed an input selection method based on combining Correlation Coefficient (CC) and Mutual Information (MI) analyses. Since DENFIS is proven to be sensitive to handle too many inputs [13], it was decided to carry out the input selection analysis started by 2 inputs and increase the number gradually to 3, 4, etc. Preliminary results showed that the model performance doesn't improve or even may deteriorate where more than 4 inputs are involved. In the adopted input selection procedure all possible combinations 2, 3, and 4 inputs are considered to identify the ones that are satisfying the selection criteria. The selected inputs for DENFIS model are R2(t-2), R7(t-1), R9(t-8), and Q(t-1), where R2, R7, and R9 refers to 2nd , 7th , and 9th rainfall stations.

4. Results and discussions

DENFIS model was calibrated with the training dataset and tested for 28 testing events by simulating discharge with D_{thr} fixed at 0.1 [14]. ARX is a linear regression model for input-output mapping where $Q(t)$ is assumed to be related to past outputs $Q(t-i)$ and inputs (rainfall) $R(t-i)$ by the following formula:

$$Q(t) = -\sum_{i=1}^{n_a} a_i Q(t-i) + \sum_{j=1}^{n_b} b_j R(t-n_k - j + 1) + e(t) \tag{2}$$

where n_a and n_b are the number of past outputs and inputs respectively, n_k is the delay associated with each input, $e(t)$ is the true error term; and a_i and b_j are model parameters to be optimized. The model performance was evaluated using coefficient of efficiency (CE), r^2, RMSE, MAE, and relative peak error (RPE) criterion. The results of DENFIS were compared against the results of ARX and HEC-HMS in order to evaluate the performance of DENFIS.

Table 1. Average CE, r^2, RMSE, and MAE values over 28 testing events resulted by DENFIS, ARX, and HEC-HMS models.

Model	CE	r^2	RMSE (m^3/s)	MAE (m^3/s)
DENFIS	0.796	0.845	6.418	3.252
ARX	0.152	0.522	10.166	7.422
HEC-HMS	0.631	0.872	7.585	4.625

The average results of DENFIS compared to ARX and HEC-HMS across 28 testing events are as shown in Table 1. As can be seen, DENFIS significantly outperformed ARX model in terms of all statistics. Moreover, DENFIS performed slightly better when compared to HEC-HMS in terms of CE, RMSE, and MAE values. However, HEC-HMS produced slightly higher r^2 compared to DENFIS. Fig. 3 compares the boxplot of RPE values across 28 testing events obtained by DENFIS and HEC-HMS models. The boxplot for ARX was unnecessary since the averaged testing results had indicated poor ARX model performance. As can be seen, DENFIS performed comparably against HEC-HMS in peak estimation but with less number of outliers.

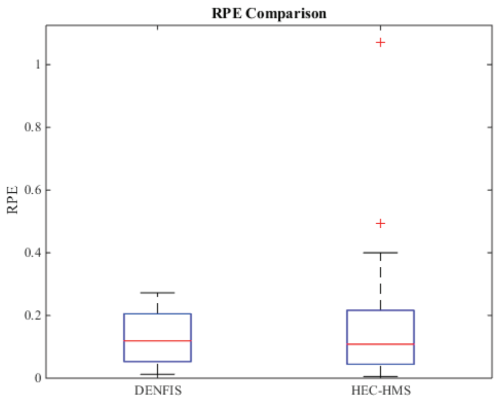


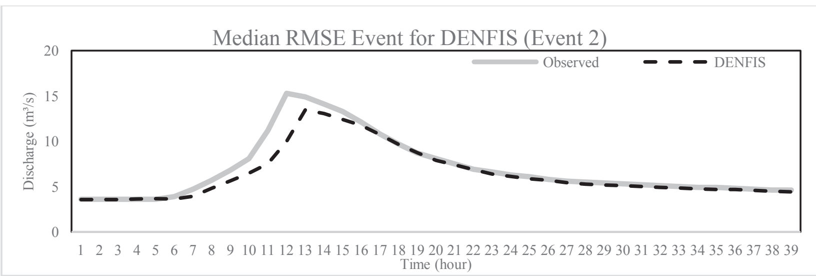
Fig. 3. Comparison of RPE values for peak discharge of 28 testing events resulted by DENFIS and HEC-HMS.

For further comparison of DENFIS and HEC-HMS models, the observed versus simulated hydrographs are shown in Fig. 4 for the events with median RMSE value. Both DENFIS and HEC-HMS models managed to follow the trend of the observed discharge; however, both models showed slight under-estimation of the peak flow. It is worth mentioning that DENFIS was able to produce comparable results against HEC-HMS by using less rainfall stations; however, DENFIS was benefited by using runoff antecedent as input unlike HEC-HMS which only uses rainfall data. In general, DENFIS was found to be a reliable alternative choice for event-based rainfall-runoff modeling.

5. Conclusion

The aim of this study was to evaluate the performance of an online learning NFS for event-based rainfall-runoff modeling in a tropical catchment. Dynamic Evolving Neural-fuzzy Inference System (DENFIS) was tested against a physically-based model and a regression model for a small urban catchment in Malaysia. Results showed that DENFIS performs significantly better than ARX model; however, when compared against HEC-HMS, DENFIS showed marginal improvements in results. It was concluded that DENFIS could be a reliable alternative for event-based rainfall-runoff modeling especially when less number of rainfall stations are planned to be utilized.

a



b

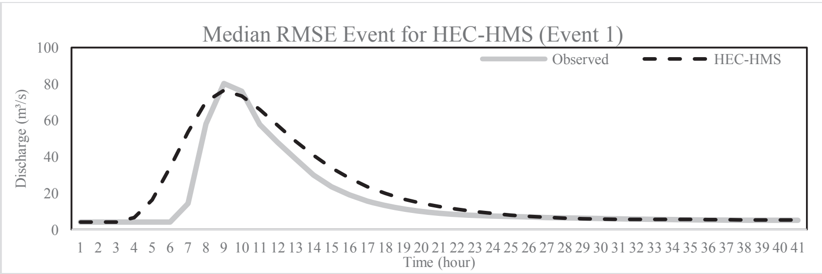


Fig. 4. Observed versus simulated hydrograph for events with median RMSE obtained by (a) DENFIS and (b) HEC-HMS.

Acknowledgements

The authors would like to thank the Fundamental Research Grant Scheme (FRGS) for providing financial support for this research. This support is granted by the Ministry of Higher Education of Malaysia.

References

[1] J.-S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," Systems, Man and Cybernetics, IEEE Transactions on, vol. 23, pp. 665-685, 1993.

[2] N. K. Kasabov and Q. Song, "DENFIS: dynamic evolving neural-fuzzy inference system and its application for time-series prediction," Fuzzy Systems, IEEE Transactions on, vol. 10, pp. 144-154, 2002.

[3] M. Desa and J. Niemczynowicz, "Spatial variability of rainfall in Kuala Lumpur, Malaysia: long and short term characteristics," Hydrological sciences journal, vol. 41, pp. 345-362, 1996.

[4] M. Desa, M. Munira, H. Akhmal, and A. Kamsiah, "Capturing extreme rainfall events in Kerayong catchment," in 10th Int. Conf. on Urban Drainage, Copenhagen Denmark, 2005, pp. 21-26.

[5] T. Piman and M. Babel, "Prediction of rainfall-runoff in an ungauged basin: Case study in the mountainous region of northern Thailand," Journal of Hydrologic Engineering, vol. 18, pp. 285-296, 2012.

[6] M. De Silva, S. Weerakoon, and S. Herath, "Modeling of Event and Continuous Flow Hydrographs with HEC–HMS: Case Study in the Kelani River Basin, Sri Lanka," Journal of Hydrologic Engineering, 2013.

[7] P. Bobál, M. Podhorányi, I. Mudroň, and M. Holubec, "Mathematical modelling of the dynamics of mountain basin snow cover in Moravian-Silesian Beskydy for operational purposes," Water Resources, vol. 42, pp. 302-312, 2015.

[8] A. Wałęga, "Application of HEC-HMS programme for the reconstruction of a flood event in an uncontrolled basin/Zastosowanie programu HEC-HMS do odtworzenia wezbrania powodziowego w zlewni niekontrolowanej," Journal of Water and Land Development, vol. 18, pp. 13-20, 2013.

[9] Y. Dinpashoh, A. Fakheri-Fard, M. Moghaddam, S. Jahanbakhsh, and M. Mirnia, "Selection of variables for the purpose of regionalization of Iran's precipitation climate using multivariate methods," Journal of Hydrology, vol. 297, pp. 109-123, 2004.

[10] A. Elshorbagy, G. Corzo, S. Srinivasulu, and D. Solomatine, "Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology-Part 2: Application," Hydrology and Earth System Sciences, vol. 14, pp. 1943-1961, 2010.

[11] J. He, C. Valeo, A. Chu, and N. F. Neumann, "Prediction of event-based stormwater runoff quantity and quality by ANNs developed using PMI-based input selection," Journal of hydrology, vol. 400, pp. 10-23, 2011.

[12] A. Talei and L. H. Chua, "Influence of lag time on event-based rainfall–runoff modeling using the data driven approach," Journal of Hydrology, vol. 438, pp. 223-233, 2012.

[13] R. Abrahart, P. E. Kneale, and L. M. See, Neural networks for hydrological modeling: CRC Press, 2004.

[14] A. Talei, L. H. C. Chua, C. Quek, and P.-E. Jansson, "Runoff forecasting using a Takagi–Sugeno neuro-fuzzy model with online learning," Journal of Hydrology, vol. 488, pp. 17-32, 2013.