

Data mining based on wavelet and decision tree for rainfall-runoff simulation

Vahid Nourani, Ali Davanlou Tajbakhsh and Amir Molajou

ABSTRACT

This study introduced a new hybrid model (Wavelet-M5 model) which combines the wavelet transforms and M5 model tree for rainfall-runoff modeling. For this purpose, the main time series were decomposed to several sub-signals by the wavelet transform, at first. Then, the obtained sub-time series were imposed as input data to M5 model tree, and finally, the related linear regressions were presented by M5 model tree. This new technique was applied on the monthly time series of Sardrud catchment and the results were also compared with other models like WANN and sole M5 model tree. The results showed that the accuracy of the proposed model is better than the previous models and also indicated the effect of data pre-processing on the performance of M5 model tree. The determination coefficient of the training stage was 0.80 and improved 31% than the M5 model tree for Sardrud catchment which is recognized as a normal watershed with a regular four seasons' pattern.

Key words | decision tree, M5 model tree, multi-linear model, rainfall-runoff modeling, wavelet transform

Vahid Nourani
Ali Davanlou Tajbakhsh
 Department of Water Resources Engineering,
 Faculty of Civil Engineering,
 University of Tabriz,
 P.O. Box 51666, Tabriz,
 Iran

Vahid Nourani
 Department of Civil Engineering,
 Near East University,
 P.O. Box 99138,
 Nicosia, North Cyprus, Mersin 10,
 Turkey

Amir Molajou (corresponding author)
 Department of Water Resources Engineering,
 Faculty of Civil Engineering,
 Iran University of Science and Technology,
 Tehran,
 Iran
 E-mail: amolajou@yahoo.com

INTRODUCTION

The rainfall-runoff (r-r) process is one of the most important components of water planning. The accurate simulation of the r-r process is a significant step in water resources management; therefore, reliable models are needed to model the intended hydrological process (Wang *et al.* 2013).

R-R process is recognized as a complex system due to the interaction of different spatio-temporal factors. Both knowledge-driven (physically based) and data-driven (system theoretic) models can simulate the r-r process through different approaches. In practice, conceptual/physical-based models require an enormous volume of environmental data and calculations to inform a detailed interaction of various physical processes controlling the hydrologic behavior of a system. When accurate predictions are more important than understanding the physics due to the complex relationships among the inputs-outputs of the r-r process, it is understood that data-driven (black box) models are reliable alternatives for physical-based models (Nourani 2017).

Black box models have recently become a popular choice based on measured observations and search to characterize the system response from those data using transfer functions (Liu & Todini 2002; Nayak *et al.* 2004; Wang *et al.* 2013; Taormina *et al.* 2015; Danandeh Mehr *et al.* 2017). However, it should be also mentioned that the data-driven models can forecast the future amounts of hydrological processes under the situation of being trained and cannot be used for extrapolation. Furthermore, the performance of data-driven techniques is highly dependent on input/output dataset quality. Thus, if the data were noisy or the correlations were weak, the ability of generalization would decrease (Danandeh Mehr *et al.* 2018; Fotovatikhah *et al.* 2018).

According to previous studies, the nature of the r-r process is non-stationary and non-linear, so linear classic statistical models such as ARIMA (Auto Regressive Integrated Moving Average) or (S)ARIMAX ((Seasonal) Auto Regressive Integrated Moving Average with eXogenous

input) are not recommended, even though they have been used before (Adamowski *et al.* 2012).

Nowadays, artificial neural network (ANN) or a hybrid model like WANN, benefiting neural networks (the combination of wavelet (WT) transform and ANN), is one of the most popular black box and non-linear models which has wide applications in different hydrological studies, including r-r modeling (Wu & Chau 2006; Chen *et al.* 2015).

In spite of the wide utilization of ANN in r-r modeling, this algorithm may exhibit defects in dealing with non-stationary r-r signals. In this situation, spatial or temporal pre-processing of data can be a necessary step to handle the non-stationary problem. In this regard, the ability of WT transform in decomposing non-stationary hydrological time series to sub-series by extracting useful information at different scales can be effective for interpreting hydrological phenomena (Nourani *et al.* 2018a, 2018b). Nourani *et al.* (2014), in a review paper, investigated the capability of hybrid WANN model in different hydrological fields in both short- and long-term scales and perceived the good efficiency of the WANN model due to the benefit of multi-resolution sub-time series as the ANN input data.

In spite of the reliable efficiency of the hybrid WANN model, in r-r modeling, some deficiencies can be attributed to WANN modeling. It is believed that the network can identify the important data; hence, the ANN users supply a large number of data as inputs. This can lead to sophisticated calculations, error, non-convergence, and over-training. Further, ANN-based models are not clear and do not help users to understand the nature of the phenomenon. The arbitrary nature of the internal representation means that there may be dramatic variations between networks of identical architecture trained on the same data. Some endeavors have been undertaken to replace understandable insights from the structure of neural networks, such as saliency analysis or the methods of recovering rules to build ANN as the 'appropriate' technique that requires better interpretability (Solomatine & Xue 2004).

Due to the scatter of input time series samples and the mentioned WANN problems, input data classification can be an appropriate tool for the reduction of chaos and complexity of datasets (Danandeh Mehr *et al.* 2017). Decision tree is one of the efficient tools of data mining in the classification, clustering, and regression issues. Decision tree, as a hierarchical

clustering method, tries to accumulate the most similar observations in each cluster (group). In other words, the observations of each group are far from other groups (the basis of division is minimizing the available entropy among the sub-groups' data). As output, decision tree provides a suitable regression model for each group. In fact, decision tree has a position between linear and non-linear models (multi-linear models) by providing piecewise linear functions and benefiting from the advantages of linear models. It should also be mentioned that the relationships presented by decision tree are simpler and more understandable by unprofessional users with regard to the complicated non-linear methods (Nourani & Molajou 2017; Nourani *et al.* 2017). M5 model tree is the most famous algorithm of decision tree which can benefit the characteristics of the classification (or clustering) and the regression methods (Quinlan 1992).

Multiple studies have been carried out regarding the efficiency of M5 model tree in simulation of hydrological processes. It was concluded that the accuracy of M5 model tree is comparable to the classic ANN model. Solomatine & Dulal (2003) investigated the performance of M5 model tree in r-r transformation and found that M5 model tree had an excellent result in prediction. Solomatine & Xue (2004) also used M5 model tree in flood forecasting problem and realized the good efficiency of this technique. Bhattacharya & Solomatine (2005) applied M5 model tree and ANN to establish a relationship between water level and discharge. It was concluded that ANN/M5 model tree was superior to the traditional model. Pal & Deswal (2009) reported the much better accuracy of M5 model tree in the case of daily reference evapotranspiration in comparison with FAO-56 Penman-Monteith equation and the calibrated Hargreaves-Samani relation. Londhe & Dixit (2012) compared the results of M5 model tree and support vector machine (SVM) technique in stream flow forecasting and reported the acceptable efficiency of M5 model tree.

In this study, as a novel strategy, a new Wavelet-M5 model was introduced and applied to Sardrud watershed. The efficiency of this technique was analyzed and compared to the other models (WANN and sole M5 model tree). Also challenged is the ability of sole M5 model tree as a pre-processing tool and the effect of WT transform on the performance of M5 model tree. It is expected that the accuracy of the proposed model will be improved due to the benefit of the

decomposed time series rather than the sole M5 model tree (Nourani *et al.* 2018a, 2018b).

MATERIALS AND METHODS

Case study: Sardrud catchment

Sardrud catchment, located in East Azarbaijan Province in north-west Iran, is one of Lake Urmia's watersheds. Zinjenab

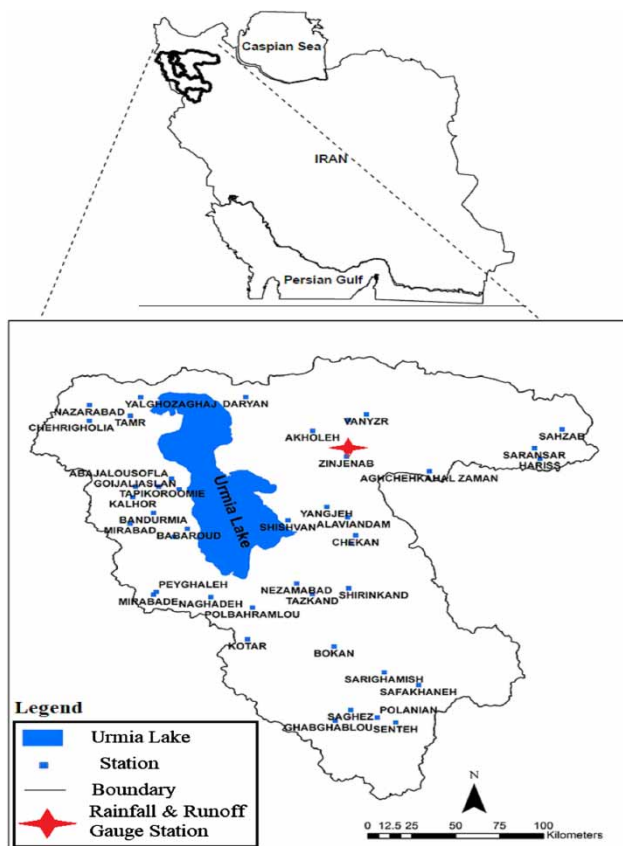


Figure 1 | The location of Sardrud watershed (Zinjenab station).

Table 1 | Statistics of the observed monthly time series for Sardrud watershed

Study area/Dataset		Rainfall time series (mm)				Runoff time series (m ³ /s)			
		Max	Min	Mean	S.D.	Max	Min	Mean	S.D.
Zinjenab	Training	160.5	0	25.204	26.262	2.358	0.001	0.294	0.35
	Verification	136	0	25.279	26.219	1.5	0.004	0.279	0.361

S.D., standard deviation.

is the main station of this watershed, which is located at geographical length 46.26° and geographical width 37.85°. Its elevation is 2,057 m above sea level. Figure 1 shows the location of Sardrud catchment and Zinjenab station. The monthly rainfall and runoff time series used is for 48 years (from 1966 until 2013, 576 months). It is noticeable that the longer period of historical records are better for modeling, but the historical records of 1966 to 2013 are used due to the quality and minimum gap (particularly as the longest available period). It should be noted that 75% of data (from year 1966 to year 2000) were dedicated to the training and the remainder (2000–2013) were used for verification.

It is strongly recommended to normalize the dataset, because the un-normalized data can have a negative effect on the performance of ANN-based models (Nourani *et al.* 2009). The statistical characteristics of Sardrud r-r time series are illustrated in Table 1. Also, the observed r-r time series are provided graphically in Figure 2.

By glancing at Table 1, it can be inferred that the standard deviations of the rainfall are scattered more than the runoff amounts. This indicates that the runoff data were closer to the mean value with a little dispersion. The difference of the data dispersion beside the correlation coefficient values for the rainfall and runoff can justify the behavior of Sardrud catchment as a normal watershed which experiences a regular precipitation pattern over a year and has a well-dominated four seasons' weather (Nourani 2017).

WT transform

The time-scale WT transform of a continuous time signal, $x(t)$, is defined as (Nourani *et al.* 2014):

$$T(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} g^* \left(\frac{t-b}{a} \right) x(t) dt \quad (1)$$

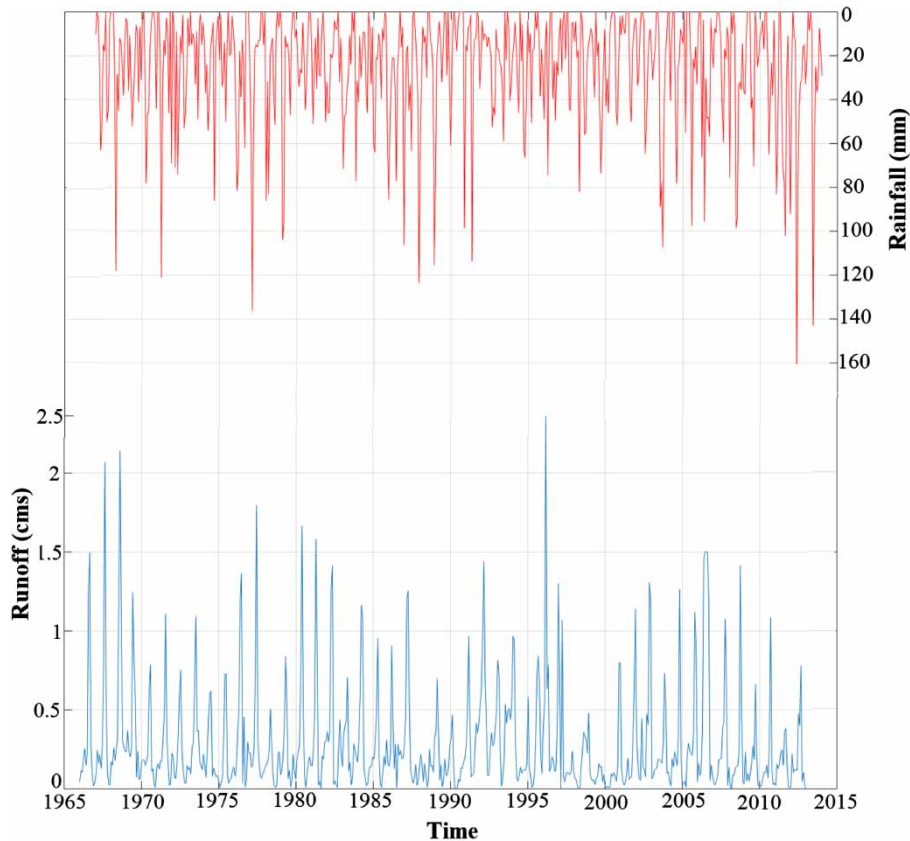


Figure 2 | Observed r-r time series (1966–2013).

where a and b define the dilation factor and temporal translation of the function $g(t)$, which permits the study of the signal around b , * corresponds to the complex conjugate, and $g(t)$ is known as the WT function or mother WT.

The important feature of the WT transform, obtained from the basic function, is providing a time-scale localization of the process. This issue would be in contrast with the classical trigonometric functions of Fourier analysis. The WT transform seeks the connections between the signal and WT function. This assessment is determined at different scales of a and locally around the time of b . The result shows a WT coefficient ($T(a,b)$) contour map known as a scalogram.

In practice, dealing with a discrete time signal to produce N^2 coefficients from a dataset of length N , based on the trapezoidal rule, using a logarithmically uniform spacing discretization of a with a correspondingly coarser resolution of the b locations, the discrete mother WT transform will

have the form of (Nourani et al. 2014):

$$g_{m,n}(t) = \frac{1}{\sqrt{a_0^m}} g\left(\frac{t - nb_0 a_0^m}{a_0^m}\right) \quad (2)$$

where a_0 , b_0 , m and n show the specified fine dilation, location parameter, integers that control the WT dilation and translation, respectively. It is noticeable that where $a_0 > 1$, a_0 is considered as 2 and where $b_0 > 0$, b_0 is usually set as 1.

M5 model tree

M5 model tree, presented by Quinlan (1992), is a kind of decision tree learning machine for regression model, meaning that it is used to forecast amounts of numerical variables. It does not attribute a constant value to the terminal node (leaf) but, instead, it fits a multi-linear regression. The M5

model tree is analogous to piecewise linear functions. The M5 model tree can learn efficiently and can control tasks with very high dimensionality. This ability has developed the popularity of the M5 model tree and caused more usage in different fields of engineering. Also, the advantage of M5 model tree over other previous linear models is that model trees are generally much smaller than regression trees and have proven more accurate in the tasks investigated.

M5 model tree partitions the data into a collection of set T and the set T is either associated with a leaf, or some test is chosen that splits T into sub-sets corresponding to the test outcomes and the same process is applied recursively to the sub-sets (Figure 3).

This process often produces over-elaborate structures, namely, over-fitting, that must be pruned. Information gain in M5 model tree is measured by reduction in standard deviation before and after the test. The first step is to compute the standard deviation of the obtained values of cases in T . Unless T contains very few cases or their values vary only slightly, T is split on the outcomes of a test. Let T_i denote a subset of cases corresponding to i_{th}

outcome of a specific test. If the deviation $sd(T_i)$ of target values of cases in T_i is considered as a measure of error, the expected reduction in error can be written as follows (Quinlan 1992):

$$SDR = sd(T) - \sum \frac{|T_i|}{|T|} sd(T_i) \quad (3)$$

Then, M5 model tree will choose one that maximizes this expected error reduction. It should be noted that the WEKA data mining software was used to extract linear regressions (M5 model tree algorithm) in this study.

WANN model

The basis of WANN model is quite similar to the ANN model, with a little difference. As the ANN is made up of three layers, the WANN also consists of a three-layer, training of error backpropagation algorithm. The input of WANN model is the decomposed r-r time series into several sub-time series by WT transform. It should be

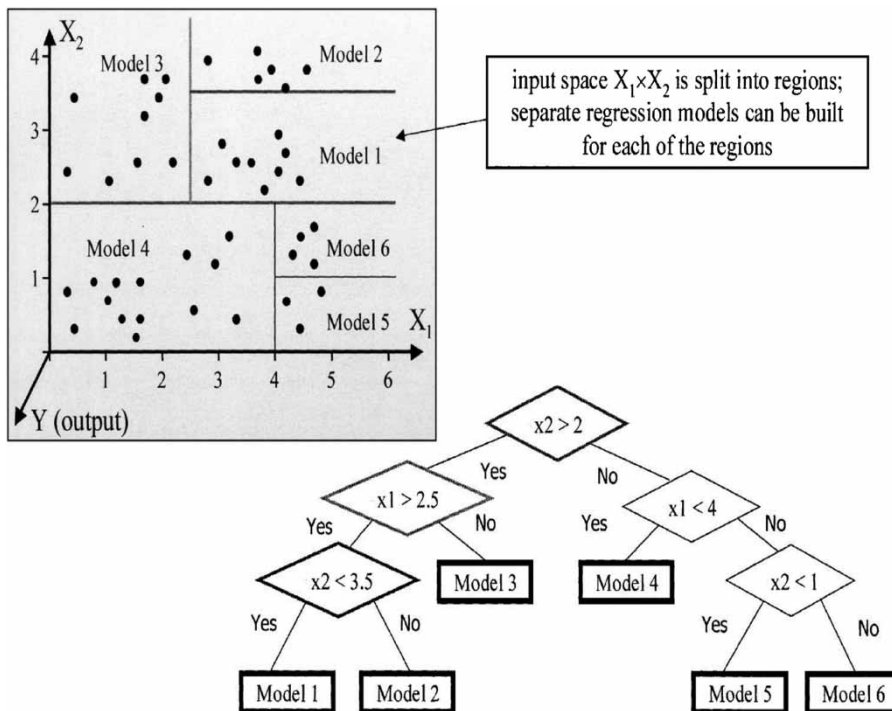


Figure 3 | An example of tree-building process in M5 model trees, dividing the domain of $X_1 \times X_2$ into six sub-domains.

noticed that WT transform deals with sub-signals in different time scales. The approximation signal ($I_a(t)$ or $Q_a(t)$), known as a large-scale sub-signal, and the d_{ith}/d_{jth} detailed signal ($I_{dith}(t)$ or $Q_{dith}(t)$) which states the short-scale sub-signals are the components of WT transform following the superposition principle (the combination of them sets the main signal) (Nourani et al. 2009). Different types of mother WT are utilized to handle the non-stationary nature of hydrological time series due to the type of studied process. According to previous studies, the db4 mother WT benefits from the adequate necessary parameters to decompose the runoff time series. Also, there are several jumps in the runoff time series because of sudden start and cessation of rainfall over the watershed. Therefore, due to the formation of db4 WT that is similar to the runoff signal, it can capture the signal features, especially peak points, efficiently and lead to comparatively good results. Due to proportional relationship between amount of rainfall and runoff, these signals were supposed to have the same seasonality level and both time series were decomposed by the same mother WT (Nourani et al. 2014).

Proposed hybrid methodology

The proposed hybrid Wavelet-M5 model combines the features of WT transform, classification (clustering), and linear regression. The architecture of the proposed model is based on linking the WT analysis to the tree-based model (M5 model tree) for r-r modeling (see Figure 4). For this purpose, at first, the main time series are decomposed to several multi-frequent sub-time series

by WT transform. Due to the relative relationship between rainfall and runoff, it was assumed that both time series included the same frequencies, so both time series were decomposed in the same level (Nourani et al. 2014). There would be many functions that can be related to the features of the main time series regarding the relationship that defines a WT function. According to previous studies, the db4 mother WT is more appropriate than other functions to simulate the r-r process since its form is more similar to the runoff signal and can capture the signal features appropriately (Nourani et al. 2014). Then, the obtained sub-time series are imposed as input data to the M5 model tree to be classified and simulated by the linear regression.

Efficiency criteria

A model for forecasting hydro-environmental time series must be evaluated in both training and verification phases. For this purpose, different statistics have been developed and used in hydro-environmental issues. Legates & McCabe (1999) showed that a hydro-environmental model can be effectively assessed via root mean square error (RMSE) and determination coefficient (DC). RMSE and DC are widely used to assess the performance of hydrological models (Sharghi et al. 2018):

$$DC = 1 - \frac{\sum_{i=1}^N (Q_i - \widehat{Q}_i)^2}{\sum_{i=1}^N (Q_i - \overline{Q_i})^2} \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Q_i - \widehat{Q}_i)^2}{N}} \quad (5)$$

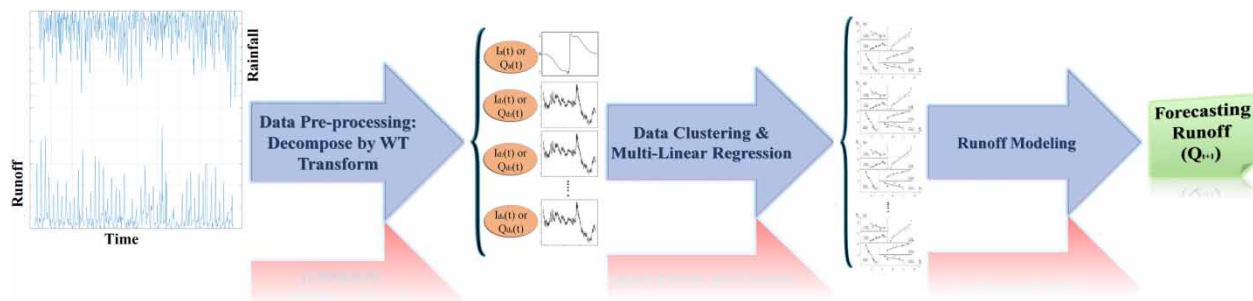


Figure 4 | The schematic structure of the proposed hybrid Wavelet-M5 model.

where DC , $RMSE$, N , Q_i , \widehat{Q}_i , \overline{Q}_i are determination coefficient, root mean square error, number of observations, observed runoff data, calculated runoff values, and mean of observed runoff data, respectively.

RESULTS AND DISCUSSION

The influence of various factors, such as precipitation, evapotranspiration, vegetation, etc., are involved in the r-r process and therefore selection of the most effective factors as input variables is an important step for r-r modeling. Previous studies have shown that r-r process is known as a Markovian or autoregressive process (Nourani 2017). Therefore, it can be stated that the effect of mentioned factors can be considered by the antecedent runoff values indirectly. Thus, the current runoff discharge would be a function of antecedent rainfall up to time step m and runoff up to time step n values as (Sharghi et al. 2018):

$$Q_t = f(I_{t-1}, I_{t-2}, \dots, I_{t-m}, Q_{t-1}, Q_{t-2}, \dots, Q_{t-n}) \quad (6)$$

Since the effects of other antecedent rainfall amounts could be implicitly considered in the previous runoff discharges, Equation (6) can be summarized as (Sharghi et al. 2018):

$$Q_t = f(I_{t-1}, Q_{t-1}, Q_{t-2}, Q_{t-3}, \dots, Q_{t-n}) \quad (7)$$

To evaluate the performance of the proposed hybrid model, the monthly r-r time series were applied and the obtained results were compared via the proposed and benchmark models (WANN and M5 model tree).

For the WANN modeling, WT transform was linked to ANN to handle the non-stationary nature of r-r time series. The r-r time series were decomposed at level 4 into five sub-time series (one approximation and four detailed sub-series) at the monthly time scales by 'db4' WT transform in order to consider the seasonal pattern of the process (it is noticeable that the WT transform was used to decompose r-r time series at level 4 in monthly modeling since $2^4 = 16$ months mode is nearly one year which is the largest period in the hydrological process (Nourani et al. 2014)). To access the best performance of ANN model in r-r simulation, the Levenberg–Marquardt scheme of back propagation algorithm was used to train ANN due to its higher convergence rate (Chau 2007; Fotovatikhah et al. 2018). Also, the sigmoid tangent activation function was used as the non-linear kernel of neural networks in this research (Nourani 2017). The network training process was stopped when the error rate was increased in the verification data. A noticeable issue, particularly in ANN/WANN modeling, which should be considered, is selecting suitable architecture of mentioned models, i.e., the number of hidden neurons and the number of iterations. The best elements were obtained by trial-and-error testing. It should be mentioned that the results of the best structures are presented in Table 2. The obtained results show the good efficiency of data pre-processing on the raw input data ($DC_{\text{training}} = 0.84$ and $DC_{\text{verification}} = 0.68$, see Table 2).

After that, it was tried to use multi-linear technique to evaluate the efficiency of the proposed model and the results were compared to WANN as a benchmark model. M5 model tree was chosen due to its ability to deal with quantitative values, classification, and producing linear regressions.

Table 2 | Results of WANN, M5 model tree, and Wavelet-M5 models for Sardrud watershed

Inputs	Output	Model	Hn	Epoch	DC		RMSE (m ³ /s)	
					Train	Verify	Train	Verify
I _t , Q _t	Q _{t+1}	WANN	9	10	0.84	0.71	0.098	0.163
		M5	–	–	0.61	0.59	0.237	0.252
		WT-M5	–	–	0.80	0.74	0.116	0.128

Hn, hidden neuron; WT-M5, Wavelet-M5.

The obtained results via M5 model tree are not comparable to the WANN model. This question asked was whether the tree-based models are capable to be considered as a pre-processing tool or not. This idea was formed to link WT transform to M5 model tree and evaluate its performance.

Similar to the WANN model, the original r-r time series were decomposed at level 4 into five sub-time series (one approximation and four detailed sub-series) at the monthly time scales via 'db4' mother WT. Then the WT-based time series were applied as inputs to M5 model tree. As was argued in the methodology, M5 model tree classifies the data samples by attributing a splitting criterion (standard deviation reduction) at the root node and branch, at first. Following this, the sample set is divided into subsets again and the splitting criterion is computed recursively for each branch. The generation of the branch is stopped when all samples at a node have the same clustering attribute at any time. Finally, a linear regression model is fitted on each subset of data samples. The results of Wavelet-M5 model are reported in Table 2. Also, to provide a view for users, the linear regressions of Wavelet-M5 are presented in the Appendix (available with the online version of this paper).

The positive effect of WT transform on the performance of M5 model tree is the first point that can be inferred from Table 2 ($(DC_{train/verify})_{WT-M5} > (DC_{train/verify})_{M5}$). The accuracy of Wavelet-M5 model is improved 31% and 25% compared to the M5 model tree in both training and verification stages, respectively. In other words, the M5 model tree did not operate as a pre-processing tool individually.

The results also show that the Wavelet-M5 model has a reliable performance and it performs as well as WANN ($(DC_{train/verify})_{WT-M5} \approx (DC_{train/verify})_{WANN}$). This can be justified by the fact that the non-linear nature of the phenomenon shows a suitable compatibility with the multi-linear model which is near and similar to its non-linear instinct. From another perspective, the behavior of the catchment and its impact on the performance of the models can be investigated. If more detailed attention is paid to Sardrud, the watershed experiences a well-dominated seasonal weather pattern. This causes its behavior to approach a semi-linear pattern. In this regard, it can be said that for catchments with a regular four-season pattern,

a simple linear or multi-linear model could be suitable for r-r simulation rather than the complex non-linear approaches.

The proximity of the DC in the training and verification steps is another point of view. As the model becomes more complicated, the DC of the training step would become better, while the DC of the verification step would still be low. Wavelet-M5 is not dependent on the number of data and is suitable for processes where a lot of historical data are not available. Since M5 model tree is the basis of the Wavelet-M5 model, all the positive features that are included in M5 model tree, such as benefiting from an understandable insight from its structure, prevention of the error magnification, the applicability of the superposition principle, and similar performance in the training and verification steps, are still true. These features could help the model to use a large number of input parameters without any change in the model accuracy, unlike the WANN model.

In addition, the computed time series via WANN, M5 model tree, and Wavelet-M5 models versus the observed time series and the scatter plots of the computed models versus the observed runoff for Sardrud catchment at the monthly scale are presented in Figure 5.

CONCLUSIONS

One of the most important features of black box models is their dependency on the case study. Although the nature of the r-r process is non-linear and non-stationary, it is expected that non-linear models would always respond better, but in some cases, the behavior of the catchment has an approximate linear pattern (experiencing well-dominated seasonal weather), thus the linear models may be effective. Also, the non-linear nature of the phenomenon can be simulated by the multi-linear model. It means instead of using the non-linear models, the multi-linear model can simulate the study process and benefit from similar accuracy. Therefore, having a view about a watershed could help hydrologists to choose a better model and not deal with the non-linear models. As a novel strategy, this study introduced a new Wavelet-M5 model. It is a combination of WT transform and M5 model tree, which has all their advantages. It was used to decompose the raw r-r time

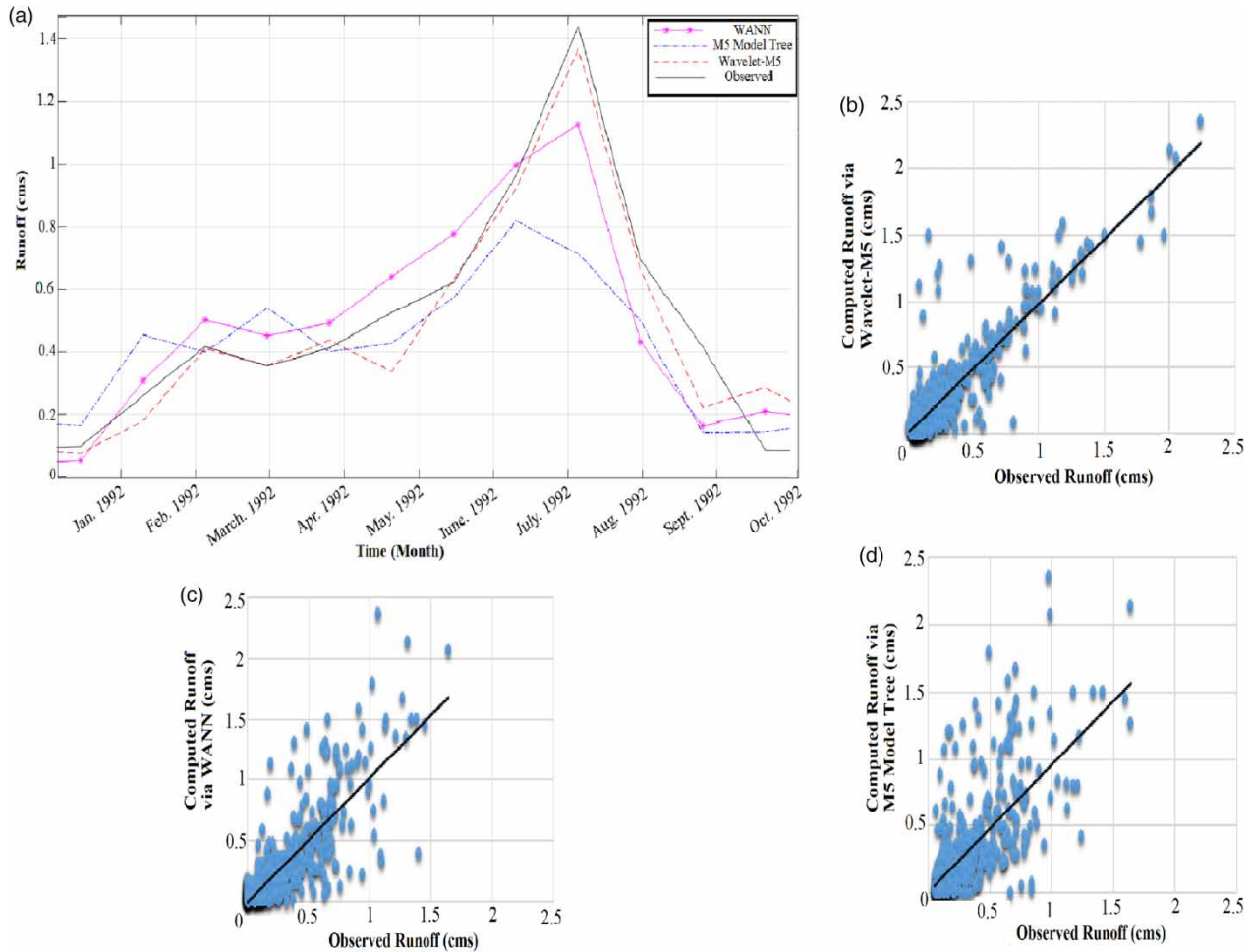


Figure 5 | Observed versus computed runoff: (a) detail and scatter plots at monthly time scale, (b) Wavelet-M5, (c) WANN, and (d) M5 model tree for Sardrud watershed.

series to several sub-time series (to eliminate the available trend by WT transform), classify the dataset, and fit a linear regression (it used multi-linear models instead of a complex non-linear regression). The accuracy of Wavelet-M5 model is improved 31% and 25% compared to M5 model tree only in the training and verification steps, respectively. Another issue challenged by the proposed hybrid Wavelet-M5 model is the ability of the individual M5 model tree as a pre-processing tool.

It is suggested that the performance of the proposed hybrid Wavelet-M5 model should be examined in different behavioral watersheds. The Wavelet-M5 model ability can also be investigated in daily time scale via different data division strategies to give a view in the presence of a low number of training data. It is also suggested that the

capability of the proposed hybrid Wavelet-M5 model is further compared with some conceptual models.

REFERENCES

- Adamowski, J., Fung Chan, H., Prasher, S. O., Ozga-Zielinski, B. & Sliusarieva, A. 2012 Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resources Research* **48** (1), <http://dx.doi.org/10.1029/2010WR009945>
- Bhattacharya, B. & Solomatine, D. P. 2005 Neural networks and M5 model trees in modelling water level-discharge relationship. *Neurocomputing* **63**, 381–396.

- Chau, K. 2007 A split-step particle swarm optimization algorithm in river stage forecasting. *Journal of Hydrology* **346** (3–4), 131–135.
- Chen, X., Chau, K. & Busari, A. 2015 A comparative study of population-based optimization algorithms for downstream river flow forecasting by a hybrid neural network model. *Engineering Applications of Artificial Intelligence* **46**, 258–268.
- Danandeh Mehr, A. D., Nourani, V., Hrnjica, B. & Molajou, A. 2017 A binary genetic programming model for teleconnection identification between global sea surface temperature and local maximum monthly rainfall events. *Journal of Hydrology* **555**, 397–406.
- Danandeh Mehr, A., Nourani, V., Karimi Khosroshahi, V. & Ghorbani, M. A. 2018 A hybrid support vector regression–firefly model for monthly rainfall forecasting. *International Journal of Environmental Science and Technology* **16** (1), 335–346. <https://doi.org/10.1007/s13762-018-1674-2>.
- Fotovatikhah, F., Herrera, M., Shamshirband, S., Chau, K.-w., Faizollahzadeh Ardabili, S. & Piran, M. J. 2018 Survey of computational intelligence as basis to big flood management: challenges, research directions and future work. *Engineering Applications of Computational Fluid Mechanics* **12** (1), 411–437.
- Legates, D. R. & McCabe, G. J. 1999 Evaluating the use of ‘goodness-of-fit’ measures in hydrologic and hydroclimatic model validation. *Water Resources Research* **35** (1), 233–241.
- Liu, Z. & Todini, E. 2002 Towards a comprehensive physically-based rainfall-runoff model. *Hydrology and Earth System Sciences* **6** (5), 859–881.
- Londhe, S. N. & Dixit, P. R. 2012 Forecasting stream flow using support vector regression and M5 model trees. *International Journal of Engineering Research and Development* **2**, 1–12.
- Nayak, P. C., Sudheer, K., Rangan, D. & Ramasastri, K. 2004 A neuro-fuzzy computing technique for modeling hydrological time series. *Journal of Hydrology* **291** (1), 52–66.
- Nourani, V. 2017 An Emotional ANN (EANN) approach to modeling rainfall-runoff process. *Journal of Hydrology* **544**, 267–277.
- Nourani, V. & Molajou, A. 2017 Application of a hybrid association rules/decision tree model for drought monitoring. *Global and Planetary Change* **159**, 37–45.
- Nourani, V., Alami, M. T. & Aminfar, M. H. 2009 A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. *Engineering Applications of Artificial Intelligence* **22** (3), 466–472.
- Nourani, V., Baghanam, A. H., Adamowski, J. & Kisi, O. 2014 Applications of hybrid wavelet – Artificial Intelligence models in hydrology: a review. *Journal of Hydrology* **514**, 358–377.
- Nourani, V., Sattari, M. T. & Molajou, A. 2017 Threshold-based hybrid data mining method for long-term maximum precipitation forecasting. *Water Resources Management* **31** (9), 2645–2658.
- Nourani, V., Danandeh Mehr, A. & Azad, N. 2018a Trend analysis of hydroclimatological variables in Urmia lake basin using hybrid wavelet Mann–Kendall and Şen tests. *Environmental Earth Sciences* **77** (5), 207.
- Nourani, V., Ghasemzadeh, M., Danandeh Mehr, A. & Sharghi, E. 2018b Investigating the effect of hydro climatological variables on Urmia Lake water level using wavelet coherence measure. *Journal of Water and Climate Change*. <https://doi.org/10.2166/wcc.2018.261>.
- Pal, M. & Deswal, S. 2009 M5 model tree based modelling of reference evapotranspiration. *Hydrological Processes* **23** (10), 1437–1443.
- Quinlan, J. R. 1992 Learning with continuous classes. In: *Proceedings of the 5th Australian Joint Conference on Artificial Intelligence*, Hobart, Tasmania.
- Sharghi, E., Nourani, V., Najafi, H. & Molajou, A. 2018 Emotional ANN (EANN) and Wavelet-ANN (WANN) approaches for Markovian and seasonal based modeling of rainfall-runoff process. *Water Resources Management* **32** (10), 3441–3456.
- Solomatine, D. P. & Dulal, K. N. 2003 Model trees as an alternative to neural networks in rainfall-runoff modelling. *Hydrological Sciences Journal* **48** (3), 399–411.
- Solomatine, D. P. & Xue, Y. 2004 M5 model trees and neural networks: application to flood forecasting in the upper reach of the Huai River in China. *Journal of Hydrologic Engineering* **9** (6), 491–501.
- Taormina, R., Chau, K.-W. & Sivakumar, B. 2015 Neural network river forecasting through baseflow separation and binary-coded swarm optimization. *Journal of Hydrology* **529**, 1788–1797.
- Wang, W.-C., Xu, D.-M., Chau, K.-W. & Chen, S. 2013 Improved annual rainfall-runoff forecasting using PSO–SVM model based on EEMD. *Journal of Hydroinformatics* **15** (4), 1377–1390.
- Wu, C. & Chau, K. 2006 A flood forecasting neural network model with genetic algorithm. *International Journal of Environment and Pollution* **28** (3–4), 261–273.

First received 24 April 2018; accepted in revised form 19 July 2018. Available online 20 August 2018