

# Artificial neural network coupled with wavelet transform for estimating snow water equivalent using passive microwave data

A B DARIANE\*, S AZIMI and A ZAKERINEJAD

*Department of Civil Engineering, K.N. Toosi University of Technology, Tehran, Iran.*

*\*Corresponding author. e-mail: borhani@kntu.ac.ir*

Snow Water Equivalent (SWE) is an important parameter in hydrologic engineering involving the stream-flow forecasting of high-elevation watersheds. In this paper, the application of classic Artificial Neural Network model (ANN) and a hybrid model combining the wavelet and ANN (WANN) is investigated in estimating the value of SWE in a mountainous basin. In addition, k-fold cross validation method is used in order to achieve a more reliable and robust model. In this regard, microwave images acquired from Spectral Sensor Microwave Imager (SSM/I) are used to estimate the SWE of Tehran sub-basins during 1992–2008 period. Also for obtaining measured SWE within the corresponding Equal-Area Scalable Earth-Grid (EASE-Grid) cell of SSM/I image, approach of Cell-SWE extraction using height–SWE relations is applied in order to reach more precise estimations. The obtained results reveal that the wavelet-ANN model significantly increases the accuracy of estimations, mainly because of using multi-scale time series as the ANN inputs. The Nash–Sutcliffe Index (NSE) for ANN and WANN models is respectively 0.09 and 0.44 which shows a firm improvement of 0.35 in NSE parameter when WANN is applied. Similar trend is observed in other parameters including RMSE where the value is 0.3 for ANN and 0.07 for WANN.

---

## 1. Introduction

Snow Water Equivalent (SWE) is identified as one of the main parameters in stream-flow forecasting of high-elevation watersheds. The problems such as limited numbers and the sparse network of snow stations and also the fact that the ground observations are cost-consuming, are main concerns for using space-borne observations to estimate snow parameters including the SWE.

Microwave images are among the most appropriate options for estimating the value of snow water equivalent. These data are not sensitive to climatic conditions such as cloudiness (Engman and Gorney 1991). In most generally applied passive microwave

(PM) snow algorithms, SWE is directly calculated via a simple regression relationship to the difference between the brightness temperature observed at 18 and 37 GHz (Chang *et al.* 1982; DeWalle and Rango 2008).

Several algorithms have been presented for estimating SWE from PM images (Foster *et al.* 1997; Tait 1998; Singh and Gan 2000). Afterwards, Koenig and Forster (2004) and Goita *et al.* (2003) evaluated the accuracy of SWE estimations from microwave images in different conditions utilizing recent algorithms. In this regard, Tedesco *et al.* (2003), Gan *et al.* (2009) and Tong *et al.* (2010) proceeded to estimate SWE using artificial neural networks and results demonstrated the superiority

**Keywords.** SWE; wavelet transform; ANN; SSM/I; k-fold cross validation; Tehran basin.

of ANN among other models. Matkan (1999) computed snow depth obtained from SSM/I data throughout Iran using a linear algorithm. Dariane et al. (2012) used SSM/I data to estimate snow water equivalent in Tehran basins.

Space-born passive microwave images have coarse resolution. Therefore, their application for estimating SWE in mountainous areas has been argued due to spatial variability of snow properties in these areas. Most PM-based research has employed EASE-Grid products having  $25 \times 25$  km fixed grid cell resolution which is very coarse for rugged surface topography of mountainous area. Because of these issues, the new GlobSnow product has omitted calculations of SWE for mountainous areas (Takala et al. 2011). Nevertheless, some researchers have suggested methods to improve the PM-based SWE estimations in these areas and benefit the available space-born images where *in situ* data are limited and sparse. For example, Li et al. (2012) showed that using the Advanced Microwave Scanning Radiometer for EOS (AMSR-E) 37 GHz Level 2A (L2A) footprints with an area of  $87.9 \text{ km}^2$  could partially overcome the resolution problem of EASE-Grid data. Tong et al. (2010) indicated that the application of ANNs for AMSR-E and SSM/I performed much better than the algorithms based on brightness temperature difference for SWE estimation on a mountainous region in Canada. Therefore, regardless of resolution issues justly mentioned by many researchers on the application of PM-based methods for estimating SWE in rugged mountainous areas, there are also ways to partially overcome the problem and to benefit the advantage of PM data in areas where *in situ* data is sparse and/or not available. This paper addresses one of these approaches through the application of wavelet transform coupled with ANN model.

Artificial Intelligence (AI) techniques are known to have great abilities in estimating nonlinear time series where they have attracted attention from various areas including hydrologic engineering. In hydrologic field, they employ available historical time series for simulating the system. In this aspect, ANN and some other methods have been in the core attention for modelling nonlinear hydrologic systems. Regardless of their relative success, ANN procedures have some shortages which limit their applications in many problems including peak flow estimations. When signal fluctuations are nonstationary, ANN-based approaches are not satisfied in precision and in this situation model performance lacks the expected quality. In order to overcome this deficiency and to raise the precision of model estimates, some hybrid models in combination with ANN are suggested (Nourani et al. 2011). A combination of wavelet transform

and ANN is one of the most common hybrid methods mentioned in the literature.

Wavelet transform method has become a popular tool in enhancing the performance of ANN models. It does this by explaining both spectral and temporal input data information in signals. In this regard, Wang and Ding (2003) applied WANN model to predict short term and long term hydrologic time series. Their results revealed that the hybrid model could increase accuracy of prediction. Partal and Cigizoglu (2008) estimated the daily suspended sediment by using a WANN model. In one step, past sediment data and in the other step, daily river flow data were applied to estimate sediment load. They reported that the WANN model has a good fit to the observed data (especially in peak values) and has better results than the classic ANN model. Adamowski and Sun (2010) also applied a coupled wavelet transform and neural network method for flow forecasting in semi-arid watershed. They used wavelet coefficients as inputs into Levenberg Marquardt artificial neural network models for flow forecasting at lead times of 1 and 3 days. Their results showed that wavelet-neural network models provide more accurate flow forecasts than the artificial neural network models. Krishna et al. (2011) used wavelet neural networks for modeling the streamflow time series in Malaprabha River (in India). They used discrete wavelet transform to decomposed time series as inputs to neural network. They compared WANN with classic ANN and autoregressive (AR) models and showed that the hybrid model is able to produce better results than either ANN or AR models.

In this article, a new application is demonstrated by using a combination of wavelet and ANN model in estimating the SWE values based on microwave images. As mentioned earlier, most previous methods for estimating SWE include linear functions. ANN method has been also applied in limited cases. Based on our information, the application of wavelet-ANN has not been yet carried out for SWE estimations.

This paper is organized as follows. The case study, ground and satellite measured data are presented in section 2. Section 3 presents the concepts and methods which are used in this study. Section 4 includes the results and discussions and finally, the last section contains the conclusions of the study.

## 2. Research area

The study area is Tehran basin that covers an area of about  $5880 \text{ km}^2$  with elevations ranging from 1236 m in the south to 4346 m in the northern region and is located in the central part of Alborz Mountains in Iran (see figure 1). This area

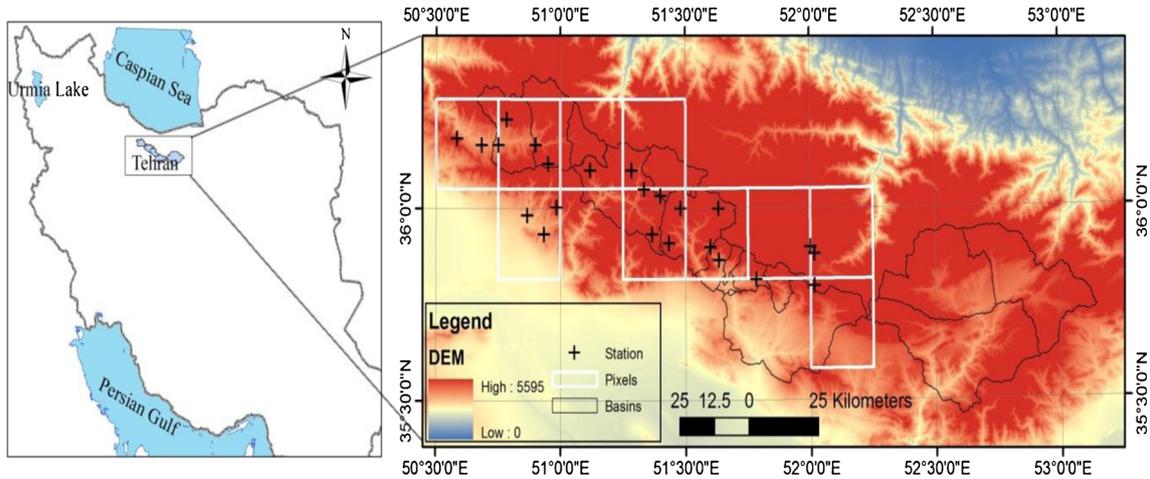


Figure 1. Location of the research area in Tehran region.

is classified with semi-humid cold climate which is affected by northern humid climate in the margin of Caspian Sea in the north, and hot and dry central plateau climate in the south (Department of Energy 2009). The area includes 23 snow measuring stations as shown in figure 1 with their corresponding SSM/I pixels.

The satellite passive microwave data are acquired by the SSM/I sensor; launched on the US Defense Meteorological Satellite Program (DMSP) in 1987. The SSM/I sensor contains seven channels that include both horizontally and vertically polarized channels at frequencies of 19, 37, and 85 GHz and a vertical polarization at 22 GHz (Armstrong *et al.* 1994–2008). The daily SSM/I data are acquired in ascending and descending mode spanning between 1992 and 2008 with the spatial resolution of 25 and 12.5 km corresponding to the channels of 19, 22, 37, and 85 GHz, respectively. The data is freely available from National Snow and Ice Data Center (NSIDC) website. In addition, a DEM from the Shuttle Radar Topography Mission (SRTM) with a spatial resolution of 90 m was used in this study.

### 3. Methods

In this section, at the beginning Cell-SWE extraction method for obtaining measured SWE within the corresponding Equal-Area Scalable Earth-Grid (EASE-Grid) cell of SSM/I image is presented. Then, the structure of artificial neural network and the concepts of wavelet transform are briefly explained. Following that, the structure of hybrid model in which wavelet transform has been combined with ANN for estimating SWE is introduced. Finally, k-fold cross validation method as a procedure to achieve more realistic model is presented.

#### 3.1 Extraction of the height–SWE relations

In most of basins in Iran, one or two snow measurements are recorded at each station during the snowy months in every year. Using these measurements, monthly average values are obtained for each station. Moreover, observed long-term average SWE values are derived in each snowy month. The height–SWE relation is then extracted for each month as:

$$\text{SWE}_m = a_m H + b_m \quad (1)$$

where  $H$  is the station height,  $a_m$  and  $b_m$  are linear regression parameters and  $\text{SWE}_m$  is the long-term average SWE for  $m$ th month.

#### 3.2 Cell-SWE extraction method

The location of snow stations whose measurements are used in this study is depicted in figure 1. As it can be found in figure 1, hydro-meteorological stations in the basin are inappropriately distributed in altitude. Therefore, in order to properly extract SWE values over EASE-Grid cells, the following methods are investigated.

- **Method 1 - simple average:** In this method, which is also a common practice,  $\text{SWE}^c$  is obtained by averaging the SWE values in each cell as:

$$\text{SWE}^c = \frac{1}{n} \sum_{i=1}^n \text{SWE}_i^c \quad (2)$$

where  $\text{SWE}_i^c$  is snow water equivalent in  $i$ th station and  $n$  is identified as number of stations in the cell.

- **Method 2 - weighted average distance:** In this approach,  $\text{SWE}^c$  in each cell is calculated based on weights extracted from the distance

between each station and the center of cell (equations 3 and 4).

$$W_i = \frac{1}{l_i} \sum_{j=1}^n \left( \frac{1}{l_j} \right) \quad (3)$$

$$\text{SWE}^c = \sum_{i=1}^n W_i \text{SWE}_i^c \quad (4)$$

where  $l_i$  is Euclidean distance from  $i$ th station to the center of cell and  $W_i$  is the weight of  $i$ th station.

- **Method 3 - normalized SWE value based on altitude and weighted average distance:** Snow and consequently SWE values in each cell depend on height; the SWE value is first adjusted to the SWE corresponding to the averaged height using the first regression parameter (slope) of the height–SWE relation, i.e.,  $a$ , and the SWE values measured at station height applying equation (5).  $\text{SWE}^c$  for each cell is then computed using equations (3) and (4).

$$\text{SWE}_n^c = \text{SWE}_{st}^c + a_m(H_{\text{avg}}^c - H_{st}^c) \quad (5)$$

where  $H_{st}^c$  is station altitude,  $H_{\text{avg}}^c$  is denoted as the averaged height of the cell  $c$ ,  $a_m$  is the slope of the height–SWE relation for  $m$ th month,  $\text{SWE}_{st}^c$  is the SWE of the corresponding station in  $m$ th month and  $\text{SWE}_n^c$  is the adjusted SWE for the  $n$ th station.

- **Method 4 - Cell-SWE extraction:** The spatial resolution of the SSM/I images is  $25 \times 25$  km while that of the SRTM DEM is  $90 \times 90$  m. Hence, the results could be improved using the SWE adjusted on the basis of DEM pixels. The SWE value of each station is corrected for DEM pixels covering one SSM/I pixel. The average values, therefore, represent the SWE of the SSM/I pixel (equations 6 and 2). The results obtained demonstrate that the proposed approach applies the height–SWE relations with higher accuracy compared to the previous methods.

$$\text{SWE}_n^c = \frac{1}{k} \sum_{PD=1}^k (\text{SWE}_{st}^c + a_m(H_{PD}^c - H_{st}^c)) \quad (6)$$

where  $H_{PD}^c$  is height in the  $PD$ th pixel of DEM that is located in the SSM/I pixel,  $k$  is the number of pixels of DEM covering the SSM/I pixel and  $\text{SWE}_n^c$  is the corrected SWE for  $n$ th station.

A comparison of results indicates that as expected, method 4 has the best performance. For example, a comparison of the last method with method 1 as shown in table 1 shows that Cell-SWE

extraction method has been able to significantly improve the correlation between SWE of ground station data and signals of SSM/I bands. Therefore, method 4 was employed for the rest of this paper.

### 3.3 Artificial neural network

Artificial intelligence methods have the ability to give better performance through handling nonlinearity and other complexities in modeling of the time series. From these categories, artificial neural networks (ANNs) have flexibilities in data modelling (Zhang and Dong 2001). ANN is reported by many authors (Tedesco et al. 2003; Gan et al. 2009) as a proper tool in estimating the value of SWE.

ANNs, also called multi-layer perceptrons (MLP), are mathematical models in order to process information and predict outputs based on network learning. Neural networks consist of an input layer, usually one or two hidden layers, and one output layer. Each layer consists of one or more neurons directionally linked with the neurons of adjacent layers. Number of neurons in the input layer is determined by the number of input parameters in the model. In the output layer, there is usually one neuron representing the output parameter. In some cases, number of neurons in the output layer could be more than one which is determined by the modeler. Number of neurons in the hidden layer is unknown and must be properly determined. A common method is by trial and error. Neurons in each layer are directionally linked with neurons of adjacent layers by weights that can be adjusted during the training process of the model. Each neuron in the hidden layer has weighted inputs and by using a transfer function over the sum of weighted inputs, it produces a weighted output that is subsequently used as the input of a neuron in the next layer. The neurons in the input layer only receive simple input values and do not have summation and/or transfer function. Some texts avoid using the term ‘layer’ for the inputs due to these characteristics. Also, the neurons in the output layer receive weighted inputs and apply summation and transfer function, but their output is not weighted. The transfer function in the output layer is usually a simple linear one. Having weights ( $w$ ) and biases ( $b$ ) and inputs ( $x$ ) of a problem, the output ( $a$ ) of each neuron using a transfer function  $f$  is defined as  $a = f(wx + b)$  (for a detailed study on this subject, see Hagan et al. 1996).

For training ANNs, datasets are introduced to the network and the output of the network is computed. Then, the mean square error is calculated over the entire dataset. In the next step, the error is back-propagated through the hidden nodes to

Table 1. Comparison of correlation coefficients between data from ground stations and SSM/I bands.

Band	H19	V19	V22	H37	V37	H85	V85
Method 4: Cell-SWE extraction	0.23	0.41	0.45	0.52	0.52	0.64	0.66
Method 1: Simple extraction	0.21	0.31	0.33	0.37	0.34	0.42	0.43

modify the weights of the inputs. This process is repeated and the weights are adjusted iteratively until the neural network converges to specified error. Back propagation algorithm is a common method for training feedforward networks.

### 3.4 Wavelet transform

Wavelet transform is relatively a new progress in the field of signals and has attracted much attention since its introduction in the early 1980s. The first aim of wavelet analysis is both to determine the frequency content of a signal and the temporal variation of this frequency content for analyzing the data. Therefore, the wavelet transform is a useful choice when signals are nonstationary (Labat 2005).

Wavelet transform analysis is a more appropriate tool than the Fourier transform in studying nonstationary signals (Partal and Kisi 2007). The main advantage of wavelets is their ability to provide information about time, location, and frequency of signals, whereas Fourier transform only provides information about the frequency (Adamowski and Chan 2011). The wavelet transforms convert a signal to a set of subsignals. These subsignals explain a better behaviour and reveal more information of the process than the original time series. So, they can help to predict with more accuracy (Remesan *et al.* 2009; Calatao 2011).

When a function satisfies the following conditions, it is a mother wavelet (Mallat 1998):

$$\int_{-\infty}^{+\infty} \varphi(t) dt = 0; \quad \int_{-\infty}^{+\infty} \frac{|\hat{\varphi}(w)|^2}{|w|} dw < \infty \quad (7)$$

where  $\varphi(t)$  is wavelet function or mother wavelet and the  $\hat{\varphi}(w)$  is Fourier transform of  $\varphi(t)$ . Other wavelets  $\varphi_{a,\tau}(t)$  can be obtained by compressing and expanding the mother wavelet:

$$\varphi_{a,\tau}(t) = a^{-1/2} \varphi\left(\frac{t-\tau}{a}\right) \quad (8)$$

where  $a$  is scale or frequency factor also called dilation factor and  $\tau$  is the time factor. The term 'scale' refers to extend or to compress the wavelet. Using small scale causes the wavelet to be compressed and

in the case of large scale, the wavelet is extended. Large scale values are not able to show the details, but small scales are applied to reveal more details.

The time-scale wavelet transform of continuous time series,  $x(t)$ , is defined as (Mallat 1998):

$$c_{\phi}^x = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \varphi^*\left(\frac{t-\tau}{a}\right) dt \quad (9)$$

where  $*$  denotes conjugate complex function.

Continuous wavelet transform (CWT) needs calculating wavelet coefficients at each scale (more work) and requires more calculation time. But as compared to CWT, the discrete wavelet transform (DWT) requires less computation time and is simpler to develop. Scales and positions of DWT are usually based on powers of two (dyadic scales and positions) and this method is more efficient than CWT for practical cases (Partal and Kisi 2007; Adamowski and Chan 2011). The discrete wavelet has the following general form (Grossman and Morlet 1984):

$$\varphi_{m,n}(t) = a_0^{-m/2} \varphi\left(\frac{t-n\tau_0 a_0^m}{a_0^m}\right) \quad (10)$$

where  $m$  and  $n$  are integers that control, respectively, the wavelet dilation (scale) and translation (time);  $a_0$  is a specified fixed dilation step greater than 1; and  $\tau_0$  is the location parameter and must be greater than zero. The most common (and simplest) choice for the parameters  $a_0$  and  $\tau_0$  is 2 and 1.

For a discrete time series  $x_i$ , where it occurs at different time  $t$ , the DWT can be defined as (Mallat 1998):

$$W_{m,n} = 2^{-m/2} \sum_{i=0}^{N-1} x_i \varphi(2^{-m}i - n) \quad (11)$$

where  $W_{m,n}$  is wavelet coefficient for the discrete wavelet of scale  $a = 2^m$  and location  $\tau = 2^m n$ .

The discrete wavelet transform is used to decompose the time series data. By this procedure, the signal is divided into two parts as approximation and detail. The process is continued by successive decomposition of the approximation. In this way, the original signal is broken down into lower resolution component.

### 3.5 Coupled wavelet and ANN (WANN)

WANN structure consists of wavelet transform and ANN model as explained in an earlier section. Wavelet-ANNs are the ANN models where the decomposed subsignals are used as the inputs. In this structure, back propagation process is used to compute weights through training the WANN model. These decomposed subsignals are obtained by discrete wavelet transform because of less computational effort than continuous wavelet transform. Wavelet-ANNs were introduced by Zhang and Benveniste (1992) as an alternative to feedforward ANNs. Wavelets can also be used within the ANN structure as substitutes to activation functions (Alexandridis and Zapanis 2013).

### 3.6 $k$ -fold cross validation

Cross validation method plays an important role in proper interpretation of the results obtained by the model. The commonly practiced method consists of dividing the dataset into two parts. Usually first 80% of the data is used for training and calibration of the model where the optimum value of adjustable parameters is determined and the remaining 20% at the end of data period is used to validate the trained model. A major setback of this method concerns the representativeness of the period selected for validation. If the validation period is a dry or wet period, or if it consists a different rainfall-runoff regime, then the validation results could be erroneous. On the other hand, employing a limited data could give optimistic or otherwise pessimistic results that are both unreal.  $k$ -fold is a method that could overcome these problems.

With  $k$ -fold method, datasets are divided into  $k$  parts. Noting that each dataset should be divided into two parts of training and validation; in each run, one fold of data is allocated for validation and  $k-1$  folds are allocated for training the model. This process is repeated  $k$  times and real error of this model is estimated by averaging the error of  $k$  runs of the model (Refaeilzadeh et al. 2009).  $k$ -fold method has considerable advantage over the common practice; for example, it makes the data play two roles of training and validation in the whole  $k$  runs. The results obtained by using  $k$ -fold may or may not be higher in evaluation parameters as compared to the classic method, but these results are real and more reliable. Therefore, the model evaluated by this method is more dependable. For estimating the actual performance of model and to obtain more reliable results,  $k$ -fold cross validation method is employed in this paper.

### 3.7 Approach

The SSM/I data acquired in descending mode is selected to estimate SWE. It should be noted that in descending mode data is recorded in the early morning which usually represents dry snow condition. Using the best Cell-SWE extraction method as described by equation (6) (method 4), the SWE value in EASE-Grid cell is calculated from the measured raw data and corresponding Digital Number (DN) values of the EASE-Grid cell extracted from seven channels. In this study, first, ANN without any data pre-processing within Tehran basin is developed to estimate the nonlinear relationship between the DN and SWE values as inputs and output of the model, respectively. In this regard, according to the measured SWE within 19 years (1990–2008, inclusive), 223 data series are obtained to establish the models. Due to rather small number of available data and for better training of the neural network, only 10% of the data is used to validate the neural network model. Thus, from 223 available data, 203 data is applied for training and the remaining 20 datasets are allocated for validation process. Also, because of small number of validation data and to achieve more reliable and robust model,  $k$ -fold resampling method is used. Moreover, the application of discrete wavelet transform in conjunction with ANN model is investigated. Data from channels 1 to 7 are decomposed by discrete wavelet transform and used as inputs of ANN. The advantage of this model over ANN lies in its ability in identifying the components of the nonstationary time series, such as SWE, by multi-level wavelet decomposition.

The models are evaluated through three well-known criteria as presented in table 2.

In forecasting models, NSE gives a better measure about the goodness of forecasts. Coefficient of determination by itself could mislead one by showing high values, where predicted flows are very different, but highly correlated with observed ones. RMSE could help to partly overcome this problem but it fails to give a firm measure of model accuracy because its value is data scale dependent. On the other hand, NSE is dimensionless and free of scale and has been widely used in hydrology and other fields of science since it was introduced by Nash and Sutcliffe (1970).

## 4. Results and discussion

In this section, first, feedforward multilayer perceptron (MLP) neural network trained by Levenberg–Marquardt's scheme of back propagation algorithm is used to estimate the SWE values. It should be mentioned that in all ANN models, a single

Table 2. Forecasting accuracy criteria.

Parameter	Formula	Definitions
Coefficient of determination ( $R^2$ )	$\frac{\left[ \sum_{i=1}^N (Q_{\text{obs}_i} - \bar{Q}_{\text{obs}}) \times (Q_{\text{pred}_i} - \bar{Q}_{\text{pred}}) \right]^2}{\sum_{i=1}^N (Q_{\text{obs}_i} - \bar{Q}_{\text{obs}})^2 \times \sum_{i=1}^N (Q_{\text{pred}_i} - \bar{Q}_{\text{pred}})^2}$	$Q_{\text{obs}_i}$ : observed, $Q_{\text{pred}_i}$ : predicted, $\bar{Q}_{\text{obs}}$ : mean observed, $\bar{Q}_{\text{pred}}$ : mean predicted, $N$ : number of data Par. Ranges:
Root mean square error (RMSE)	$\sqrt{\frac{\sum_{i=1}^N (Q_{\text{obs}_i} - Q_{\text{pred}_i})^2}{N - 1}}$	$(0 \leq R^2 \leq 1)$ $(0 \leq \text{RMSE} \leq \infty)$ $(-\infty \leq \text{NSE} \leq 1)$
Nash–Sutcliffe model efficiency (NSE)	$1 - \frac{\sum_{i=1}^N (Q_{\text{obs}_i} - Q_{\text{pred}_i})^2}{\sum_{i=1}^N (Q_{\text{obs}_i} - \bar{Q}_{\text{obs}})^2}$	

hidden layer is assumed and the number of neurons in this layer is selected by trial and error. Also, for better performance of the neural network model, available data are normalized between zero and one. So, SWE is presented in dimensionless form in figures.

In this case, using k-fold method, datasets are divided into 11 parts (folds) and then in each run, one fold of data is set aside for validation and the remaining 10 parts are used for training the model. It was mentioned earlier, data from channels 1 to 7 (seven inputs) are used as input variables to estimate the SWE as the model output. The results obtained by ANN model using k-fold resampling method are presented in table 3.

According to table 3, the model in stage 5 has the weakest results with NSE value equal to  $-1.77$ . This means that probably the dataset used for validation in this stage has a different trend than the rest of data that are used in calibration of the model. Also, the model in stage 11 shows the strongest performance with an NSE value of 0.55. This is also probably an indication that the validation data in this stage is in accordance with the main trend of the training data. It is evident from these results that if a classic method was used we could have made erroneous conclusions about the model performance. The average performance of the model has an NSE value of about 0.10. The model in stage 6 has a close performance to the average NSE which is selected for further analysis. So, this part of data (inclusive 2002–2003) is used as the validation period in figure 2. It is necessary to mention that because of using normalized data, the RMSE values are dimensionless throughout this paper.

Figure 2 shows the estimated and observed SWE and scatter plot of selected ANN model for validation period. The solid line in the left plot shows the perfect prediction where data calculated by the model perfectly matches the observed ones.

It is evident from the data scatterings in the left plot that the model predictions are insensitive to variations of the observed data. This explains the point that due to low NSE value of 0.10, which is close to zero, the model performance is not much better than the observed average. Obviously, this model would not be suitable for SWE estimates.

In the next step, a hybrid model using the pre-processed data, generated by the wavelet method, is used as the input to the ANN model in order to improve the results. In other words, wavelet decomposition helps obtaining detailed and fine features from the time series signals that enter as the inputs to the ANN model. These subsignals explain a better behaviour of the process and reveal more characteristics of the input data by which the ANN receives better training. In this article, the wavelet-ANN (WANN) model is established by applying subsignals computed through discrete wavelet transform (DWT) on original signals. Signals are the data related to channels 1 to 7 of SSM/I sensor.

Through wavelet decomposition method, one approximation subsignal of the original signal and several detail subsignals (equal to number of decomposition stages) are obtained and these subsignals of discrete wavelet transform show variations on different scales. In this application, coiflet 1 was applied as mother wavelet by trial and error and 2 decomposition levels are selected for decomposing time series. Time series with scales  $2^1$ ,  $2^2$ , and approximate and original signals of data related to channel 1 are presented in figure 3. With these decomposed sub-signals as the input of ANN model, 21 inputs (one approximate and two details for each signal), related to 1990–2008 duration, are applied to estimate the SWE values.

Once again the k-fold resampling method is used following all the previously explained steps. The results obtained by WANN model as presented in

Table 3. Results of ANN model using k-fold resampling method.

Stage	Training period	Validation period	Validation period		
			NSE	( $R^2$ )	RMSE
1	The remaining	Part 1 (Twenty-data)	0.18	0.28	0.33
2		Part 2 (Twenty-data)	0.05	0.06	0.35
3		Part 3 (Twenty-data)	0.18	0.26	0.22
4		Part 4 (Twenty-data)	0.36	0.44	0.25
5		Part 5 (Twenty-data)	-1.77	0.68	0.24
6		Part 6 (Twenty-data)	<b>0.09</b>	<b>0.24</b>	<b>0.30</b>
7		Part 7 (Twenty-data)	0.46	0.65	0.26
8		Part 8 (Twenty-data)	0.53	0.58	0.20
9		Part 9 (Twenty-data)	0.19	0.24	0.24
10		Part 10 (Twenty-data)	0.35	0.48	0.2
11		Part 11 (Twenty-three data)	0.55	0.6	0.14
Mean:			0.10	0.41	0.25
Min:			-1.77	0.06	0.14
Max:			0.55	0.68	0.35

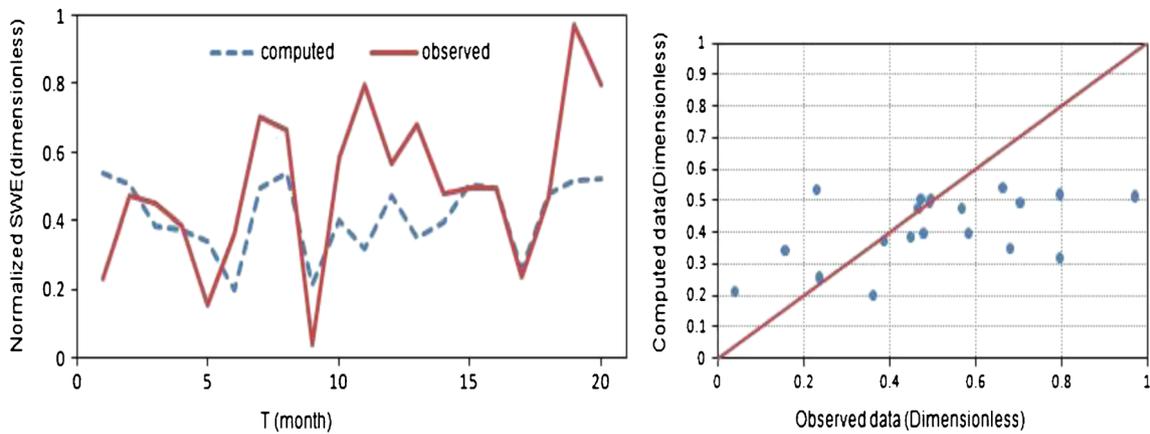


Figure 2. The estimated and observed SWE and scatter plot of ANN model for validation period.

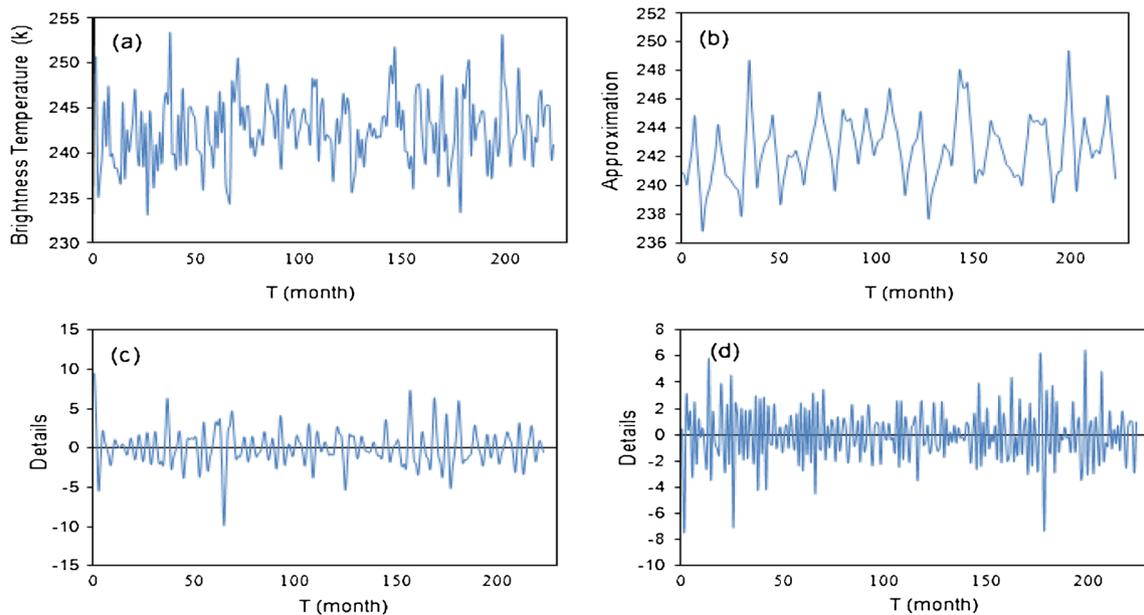


Figure 3. (a) Original input signal (channel 1); (b) approximation subsignal of input signal (level 2); (c) detail subsignal of input signal (level 2); and (d) detail subsignal of input signal (level 1).

table 4 show a substantial improvement over the ANN model.

Table 4 shows that the model in stage 9 has the weakest performance with an NSE value equal to 0.21. Also, the WANN model developed in stage 11 shows the best performance with NSE equal to 0.68. The average NSE obtained for all k-folds is about 0.43 which is a significant result for passive microwave application in mountainous areas. We observe that the model in stage 5 has the closest performance to the average of all runs which is selected for further analysis. In addition, this part of dataset (inclusive 2000–2002) is used as the validation period in figure 4.

Figure 4 shows the estimated and observed SWE and scatter plot of selected WANN model for validation period. The solid line in the left plot shows the place of perfect prediction.

A comparison of results obtained by the classic ANN model with those of the hybrid WANN model reveals improvement in results by using hybrid model. The average NSE of ANN model is 0.10 (table 3) whereas for the WANN model, mean NSE is 0.43 (table 4), indicating a substantial improvement. Other parameters including average  $R^2$  and RMSE for ANN are 0.41 and 0.25, respectively, while they are 0.50 and 0.15 for WANN model, also showing a similar trend. The improvements in the maximum and minimum of parameters are more noticeable. Minimum NSE in ANN model is  $-1.77$  which is increased to 0.21 in the WANN model. The real value of hybrid model is more appreciated by comparing the plots in figures 2 and 4. Comparison of time series plots and the scatter plots shows that the SWE values estimated by the WANN model are more precise than those found by the ANN.

Table 4. Results of hybrid WANN model using k-fold resampling method.

Stage	Training period	Validation period	Validation period		
			NSE	$R^2$	RMSE
1	The remaining	Part 1 (Twenty-data)	0.34	0.35	0.21
2		Part 2 (Twenty-data)	0.32	0.32	0.21
3		Part 3 (Twenty-data)	0.25	0.28	0.15
4		Part 4 (Twenty-data)	0.5	0.6	0.16
5		Part 5 (Twenty-data)	<b>0.44</b>	<b>0.56</b>	<b>0.07</b>
6		Part 6 (Twenty-data)	0.36	0.56	0.18
7		Part 7 (Twenty-data)	0.54	0.67	0.17
8		Part 8 (Twenty-data)	0.64	0.70	0.13
9		Part 9 (Twenty-data)	0.21	0.30	0.17
10		Part 10 (Twenty-data)	0.51	0.53	0.13
11		Part 11 (Twenty-three data)	0.68	0.68	0.09
Mean:			0.43	0.50	0.15
Min:			0.21	0.28	0.07
Max:			0.68	0.70	0.21

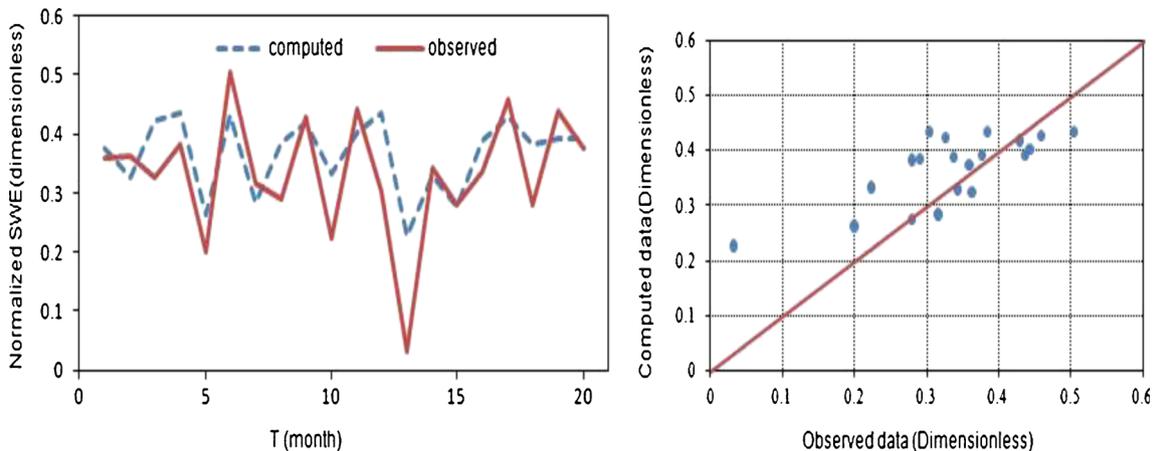


Figure 4. The estimated and observed SWE and scatter plot of WANN model for validation period.

## 5. Conclusion

In this paper, the application of classic ANN and a hybrid model combining the wavelet and ANN (WANN) was investigated in estimating the value of SWE in a mountainous basin. Microwave images acquired from Spectral Sensor Microwave Imager (SSM/I) were used to estimate the SWE of Tehran sub-basins during 1992–2008 period. Also for obtaining measured SWE within the corresponding Equal-Area Scalable Earth-Grid cell of SSM/I image, approach of Cell-SWE extraction was successfully used. It was found that the application of neural networks in SSM/I images helps in better estimations of SWE values. Moreover, wavelet transform when combined with ANN was able to substantially enhance the performance of the model. A comparison of results obtained by the classic ANN model without any data pre-processing with those of the hybrid WANN model that employs wavelet transform to pre-process the input signals revealed the true value of wavelet transform. The average NSE of ANN model was found to be 0.10, whereas for the WANN model, mean NSE was estimated at 0.43, indicating a substantial improvement. In addition, comparison of time series plots and the scatter plots showed that the SWE values estimated by the WANN model are more precise than those found by the ANN.

## Acknowledgments

Authors would like to thank NSIDC for providing SSM/I data and the Iranian Water Resource Management Organization for making available ground data needed in this project.

## References

- Adamowski J and Sun K 2010 Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds; *J. Hydrol.* **390** 85–91.
- Adamowski J and Chan H F 2011 A wavelet neural network conjunction model for groundwater level forecasting; *J. Hydrol.* **407** 28–40.
- Alexandridis A K and Zapranis A D 2013 Wavelet neural networks: A practical guide; *Neural Netw.* **42** 1–27.
- Armstrong R L, Knowles K W, Brodzik M J and Hardman M A 1994 Updated current year. DMSP SSM/I Pathfinder Daily EASE-Grid Brightness Temperatures [1995–2008]; National Snow and Ice Data Center, Digital media, Boulder, Colorado, USA.
- Calatao J P S 2011 Hybrid wavelet-PSO-ANFIS approach for short-term wind power forecasting in Portugal; *IEEE Trans. Sust. Ener.* **2**(1) 50–59.
- Chang A T C, Foster J L, Hall D K, Rango A and Hartline B K 1982 Snow water equivalent estimation by microwave radiometry; *Cold Reg. Sci. Technol.* **5** 259–267.
- Dariane A B, Zakeri Nejad A R and Dehghani M 2012 Estimation of snow water equivalent using SSM/I passive microwave data in Tehran; *9th International Congress on Civil Engineering*, Isfahan University of Technology, May 8–10.
- Department of Energy, Iran Water Resources Management Company, Tehran Regional Water Company 2009 Organize the surface waters south of Tehran (qualitative and quantitative studies) Volume I: Report of Weather; report code: TWM/MS-02 (in Persian).
- DeWalle D R and Rango A 2008 *Principle of snow hydrology*; England, Cambridge University, 410p.
- Engman E T and Gorney R J 1991 *Remote sensing in hydrology*; London, Chapman and Hall.
- Foster J L, Chang A T C and Hall D K 1997 Comparison of snow mass estimates from a prototype passive microwave snow algorithm, a revised algorithm and snow depth climatology; *Remote Sens. Environ.* **62** 132–142.
- Gan T Y, Kalinga O and Singh P 2009 Comparison of snow water equivalent retrieved from SSM/I passive microwave data using artificial neural network, projection pursuit and nonlinear regressions; *Remote Sens. Environ.* **113** 919–927.
- Goita K, Walker A E and Goodison B E 2003 Algorithm development for the estimation of snow water equivalent in the boreal forest using passive microwave data; *Int. J. Remote Sens.* **24** 1097–1102.
- Grossman A and Morlet J 1984 Decompositions of hardy functions into square integrable wavelets of constant shape; *SIAM J. Math. Anal.* **5** 723–736.
- Hagan M T, Demuth H B and Beale M H 1996 *Neural Network Design*; Boston, PWS Publishing.
- Koenig L S and Forster R R 2004 Evaluation of passive microwave snow water equivalent algorithms in the depth hoar-dominated snowpack of the Kuparuk River Watershed, Alaska, USA; *Remote Sens. Environ.* **93** 511–527.
- Krishna B, Rao Y and Nayak P 2011 Time Series Modeling of river flow using wavelet neural networks; *JWARP* **3**(1) 50–59.
- Labat D 2005 Recent advances in wavelet analysis: Part 1 – a review of concepts; *J. Hydrol.* **314** 275–288.
- Li D, Durand M and Margulis S A 2012 Potential for hydrologic characterization of deep mountain snowpack via passive microwave remote sensing in the Kern River basin, Sierra Nevada, USA; *Remote Sens. Environ.* **125** 34–48.
- Mallat S G 1998 *A wavelet tour of signal processing*; 2nd edn, San Diego, Academic Press.
- Matkan A 1999 *Passive microwave monitoring of snow cover and rainfall over Iran using DMSP: F-11 SSM/I data*; Bristol University.
- Nash J E and Sutcliffe J V 1970 River flow forecasting through conceptual models. Part I – A discussion of principles; *J. Hydrol.* **10**(3) 282–290.
- Nourani V, Kisi O and Komasi M 2011 Two hybrid Artificial Intelligence approaches for modeling rainfall–runoff process; *J. Hydrol.* **402** 41–59.
- Partal T and Cigizoglu H K 2008 Estimation and forecasting of daily suspended sediment data using wavelet–neural networks; *J. Hydrol.* **358** 317–331.
- Partal T and Kisi O 2007 Wavelet and neuro-fuzzy conjunction model for precipitation forecasting; *J. Hydrol.* **342** 199–212.
- Refaeilzadeh P, Tang L and Liu H 2009 Cross-validation; In: *Encyclopedia of Database Systems* (eds) Liu L and Ozsu T, Springer, pp. 532–538.
- Remesan R, Shamim M A, Han D and Mathew J 2009 Runoff prediction using an integrated hybrid modelling scheme; *J. Hydrol.* **372** 48–60.

- Singh P R and Gan T Y 2000 Retrieval of snow water equivalent using passive microwave brightness temperature data; *Remote Sens. Environ.* **74** 275–286.
- Tait A B 1998 Estimation of snow water equivalent using passive microwave radiation data; *Remote Sens. Environ.* **64** 286–291.
- Takala M, Luojus K, Pulliainen J, Derksen C, Lemmetyinen J, Kärnä J P and Koskinen J 2011 Estimating northern hemisphere snow water equivalent for climate research through assimilation of spaceborne radiometer data and ground-based measurements; *Remote Sens. Environ.* **115** 3517–3529, doi: [10.1016/j.rse.2011.08.014](https://doi.org/10.1016/j.rse.2011.08.014).
- Tedesco M, Pulliainen J, Takala M, Halikainen M and Pampaloni P 2003 Artificial neural network-based techniques for the retrieval of SWE and snow depth from SSM/I data; *Remote Sens. Environ.* **90** 76–85.
- Tong J, Dery S J, Jackson P L and Derksen C 2010 Testing snow water equivalent retrieval algorithms for passive microwave remote sensing in an alpine watershed of western Canada; *Can. J. Remote Sens.* **36** 74–86.
- Wang W and Ding S 2003 Wavelet network model and its application to the prediction of hydrology; *Nature Sci.* **1(1)** 67–71.
- Zhang Q and Benveniste A 1992 Wavelet networks; *IEEE T. Neural Networks* **3(6)** 889–898.
- Zhang B L and Dong Z Y 2001 An adaptive neural-wavelet model for short term load forecasting; *Electer. Pow. Syst. Res.* **59** 121–129.

*MS received 25 February 2013; revised 16 May 2014; accepted 17 May 2014*