A wavelet-SARIMA-ANN hybrid model for precipitation forecasting

Maryam SHAFAEI[1) ABCDEF], **Jan ADAMOWSKI**[2) ABCDEF],
Ahmad FAKHERI-FARD[1) ABCDEF], **Yagob DINPASHOH**[1) ABCDEF],
Kazimierz ADAMOWSKI[3) ABCDEF]

[1)] University of Tabriz, Department of Water Engineering, Iran; e-mail: maryamshafaei65@yahoo.com, fakheri@tabrizu.ac.ir, dinpashoh@yahoo.com
[2)] McGill University, Faculty of Agricultural and Environmental Sciences, Department of Bioresource Engineering, Quebec, Canada, H9X 3V9; e-mail: jan.adamowski@mcgill.ca
[3)] University of Ottawa, Department of Civil Engineering, Canada; e-mail: kadamowski@hotmail.com

Abstract

Given its importance in water resources management, particularly in terms of minimizing flood or drought hazards, precipitation forecasting has seen a wide variety of approaches tested. As monthly precipitation time series have nonlinear features and multiple time scales, wavelet, seasonal auto regressive integrated moving average (SARIMA) and hybrid artificial neural network (ANN) methods were tested for their ability to accurately predict monthly precipitation. A 40-year (1970–2009) precipitation time series from Iran's Nahavand meteorological station (34°12'N lat., 48°22'E long.) was decomposed into one low frequency subseries and several high frequency sub-series by wavelet transform. The low frequency sub-series were predicted with a SARIMA model, while high frequency subseries were predicted with an ANN. Finally, the predicted subseries were reconstructed to predict the precipitation of future single months. Comparing model-generated values with observed data, the wavelet-SARIMA-ANN model was seen to outperform wavelet-ANN and wavelet-SARIMA models in terms of precipitation forecasting accuracy.

Key words: *artificial neural network (ANN), precipitation forecasting, seasonal auto regressive integrated moving average (SARIMA), water resources management, wavelet*

INTRODUCTION

Precipitation is a major hydrologic variable, especially in the context of climate change, and as such its prediction is important in water resources engineering, planning and management [ADAMOWSKI *et al.* 2010; HALBE *et al.* 2013; SAADAT *et al.* 2011]. The spatiotemporally complex atmospheric process of precipitation is difficult to predict given the apparent stochastic properties of precipitation series. To set water resource planning goals, long-term precipitation series are required and many efforts have been made to find the most suitable procedure for predicting precipitation. A number of classic time series models, including auto-regressive integrated moving average (ARIMA) models, have been developed for hydrological time series prediction [SALAS *et al.* 1980]. Assuming data to be stationary, these essentially linear models have a restricted ability to capture non-stationarity and non-linearity in hydrological data. Other precipitation prediction methods include pairing physical, marine, meteorological or satellite data with

a prediction model, or employing data handling tools such as artificial neural networks (ANNs) or fuzzy logic. Of the several numeral weather forecasting models recently developed, most have been limited to short term predictions [TANTANEE *et al.* 2005].

As self-learning and self-adaptive function approximators, ANNs have shown a great capacity to model and predict non-linear hydrologic time series [ADAMOWSKI *et al.* 2012a; 2013; BELAYNEH *et al.* 2014; GOYAL *et al.* 2014; HAIDARY *et al.* 2013; KARRAN *et al.* 2014; RATHINASAMY *et al.* 2015; TIWARI, ADAMOWSKI 2014]. In spite of the frequently successful use of ANNs in modeling hydrologic time series, on occasion they perform poorly, particularly when signal vacillations are highly non-stationary or physical hydrologic process occur under a large range of time scales (*e.g.*, 1 day to several decades). Under such conditions, ANNs may not be capable of overcoming the limitations imposed by the data's non--stationarity without pre-processing of the input and/or output data [CANNAS *et al.* 2006]. Wavelet transform is a powerful tool to apply when confronted with non-stationary time series as it is uniquely capable of discerning unusual events (outliers) through localized time and frequency analysis [ADAMOWSKI, PROKOPH 2013; AKANSU *et al.* 2010; ARAGHI *et al.* 2015; CAMPISI *et al.* 2012; NOURANI *et al.* 2014; RATHINASAMY *et al.* 2014; SELESNICK *et al.* 2005; TIWARI, ADAMOWSKI 2015].

Wavelet transforms address the fundamental weakness of Fourier analysis, namely that the Fourier spectrum contains only globally averaged data. Data pre-processing can be performed by time series decomposition into subseries using wavelet transforms. The wavelet decomposition of non-stationary time series into distinctive scales provides an explanation of the series' structure and extracts meaningful information about its history, using several coefficients [NASON, VON SACHS 1999]. In wavelet transform, time series data are decomposed into approximations and detailed subseries; approximations are the high-scale, low frequency constituents of the signal, and the details are the low-scale and high frequency constituents [ADAMOWSKI *et al.* 2012b]. For example, wavelet analysis was employed for the decomposition of interdecadal and interannual portions of northern China's rainy season (summer monsoon) precipitation time series [LU 2002]. TANTANEE *et al.* [2005], employing a wavelet-ARIMA procedure to predict precipitation in northeastern Thailand, found it to outperform an ARIMA model operating alone.

Applying wavelet-gene-expression programming (WGEP) and wavelet-neuro-fuzzy (WNF) models to predict daily precipitation in western Turkey, KISI and SHIRI [2011] found the new hybrid WGEP model to be more effective than the WNF models. Similarly, using a combined wavelet-support vector machine model for daily precipitation prediction at two meteorological stations in western Turkey, KISI and CIMEN [2012] found the combined model to exhibit greater prediction accuracy than the support vector machine model alone. These studies demonstrate the efficiency and accuracy of wavelet transform as a tool for precisely locating irregularly distributed multi-scale spatiotemporal particularities in climate elements.

The main aim of the present study was to predict precipitation at the Nahavand meteorological station in Iran using a wavelet-SARIMA-ANN (WSA) model. The main advantage of wavelet transform is that it allows one to study different treatments independently on distinct time scales. The Nahavand precipitation series data was only decomposed to 1, 2 and 3 levels as further decomposition levels did not improve prediction accuracy. Given ANNs ability to predict non--stationary and non-linear time series, the detail subseries were predicted using an ANN. However, the approximations were predicted using a SARIMA model, given this model's particular suitability for forecasting stationary time series. The predicted approximation and detail subseries were subjected to an inverse wavelet transform to reconstruct the signal. The precipitation forecasting accuracy of the wavelet-SARIMA-ANN model was compared to that of wavelet-ANN and wavelet-SARIMA models. The sensitivity of pre-processing to wavelet type (coif1, db3, or dmey) and decomposition level were also examined.

WAVELET TRANSFORM

Mathematical functions which serve in the analysis of non-stationary time-series, wavelet transforms allow one to decompose time series into low frequency and high frequency information, thereby exposing trends, break-down points, and discontinuities in the data that other signal analysis methods might miss [ADAMOWSKI *et al.* 2009; KIM, VALDES 2003; NALLEY *et al.* 2012; 2013]. Another advantage of wavelet transforms is the flexibility of choice in selecting a mother wavelet according to the properties of the time series [ADAMOWSKI, CHAN 2011; ARAGHI *et al.* 2014; PINGALE *et al.* 2014]. Continuous wavelet analysis (CWT) represents the sum over all time of the signal, multiplied by scale and shifted versions of the mother wavelet analysis ψ [KIM, VALDES 2003]:

$$W(\tau,s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{+\infty} x(t)\psi * \left(\frac{t-\tau}{s}\right) \mathrm{d}t \qquad (1)$$

where, s is the scale parameter, t is time, and τ is the shift parameter [KIM, VALDES 2003]. Each scale corresponds to the width of the wavelet. While a CWT is useful in processing various images and signals, it is seldom used for prediction as its calculations are intricate and lengthy in terms of time. As an alternative, in prediction applications, the discrete wavelet transform (DWT) is applied, due to its simplicity and shorter calculation time. The scales and shifts of the DWT are usually based on powers of two (dyadic scales and shifts). This is obtained by altering the wavelet transform:

$$\psi(m)_{j,m} = \frac{1}{\sqrt{|s_0^j|}} \sum_k \psi\left(\frac{k - m\tau_0 s_0^j}{s_0^j}\right) x(k) \qquad (2)$$

where, j and m are integers which control the scale and shift, respectively, $s_0 > 1$ is a fixed expansion step, and τ_0 is a shift parameter based on the afore-mentioned expansion step. The impact of discretizing the wavelet transform is that the time-space scale is sampled at discrete levels. The DWT functions like a pair of high-pass and low-pass filters. The time series is decomposed into one comprising its trend (the approximation) and one comprising the high frequencies and the fast events (the details) [ADAMOWSKI, SUN 2010]. In the present study, the detail coefficients and approximation sub-time series were obtained using eq. (2).

MATERIALS AND METHODS

ARTIFICIAL NEURAL NETWORKS

In most cases, an ANN is an adaptive system that alters its structure based on external or internal data passing through its network during the learning stage. An ANN model is therefore a nonlinear statistical data tool, which can thus be applied to model relationships between inputs and outputs. Although several ANN models have been suggested, the most popular for time series prediction are multi-layer feed-forward networks [SANTOS *et al.* 2003]. Application of an ANN includes three basic steps: network architecture determination, network learning, and network verification. The architecture is determined by the number of layers, number of neurons per layer, type of connection between layers (activation functions), and the type of network. One of the most commonly used algorithms for ANN learning is the error back-propagation algorithm, which is applied in the present paper. Finally, in the verification process, the network is used to simulate a data set which was not used during the learning process. The results are then compared with the observed data in order to verify whether the trained network is able to generalize the results obtained during the learning process [SANTOS *et al.* 2009].

SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE

Seasonal autoregressive integrated moving average (SARIMA) is a popular linear model for predicting seasonal time series. A time series $\{Z_t | 1, 2, \ldots, k\}$ is created by the SARIMA process of Box and Jenkins time series modeling [BOX, JENKINS 1976] if:

$$\phi_p(B)\varphi_p B^s (1-B)^d (1-B^s)^D Z_t = \theta_q(B)\Theta_Q(B^s) \qquad (3)$$

Where p, d, q, P, D, Q are integers, s is the season length:

$$\phi_p(B) = 1 - \phi_1(B) - \phi_2 B^2 - \cdots - \phi_p B^p \qquad (4)$$

$$\varphi_P(B^s) = 1 - \varphi_s B^s - \phi_{2s} B^{2s} - \cdots - \phi_{Ps} B^{Ps} \qquad (5)$$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q \qquad (6)$$

$$\Theta_Q(B^s) = 1 - \Theta_s B^s - \Theta_{2s} B^{2s} - \cdots - \Theta_{Qs} B^{Qs} \qquad (7)$$

As polynomials in B, where B is the backwards transfer factor, ε_t is the estimated residual at time t, d is the number of normal difference, D is the number of seasonal differences, Z_t indicates the observed value at time t, $t = 1, 2, \ldots, k$. ε_t is independently and definitely distributed as a normal random variable with mean 0 and constant variance σ^2. The roots of $\phi_p(Z) = 0$ and $\theta_q(Z) = 0$ are all situated outside the unit circle. Fitting a SARIMA model to data entails a four-step iterative process: (i) discerning the SARIMA (p, d, q) (P, D, Q) structure; (ii) estimating unknown parameters; (iii) carrying out goodness of-fit tests on the estimated residuals; and (iv) predicting future values based on the known data. The fitting of SARIMA models requires the use of autocorrelation function (ACF) charts.

MODEL DEVELOPMENT

In this study some irregular mother wavelets (*e.g.*, coif1, db3, dmey) were applied for decomposition of the time series. To predict precipitation one month ahead of time, the precipitation time series was decomposed into one, two and three decomposition levels using coif1, db3 or dmey mother wavelets through code written in Matlab Software. Increasing the number of decomposition levels beyond three, and therefore the number of sub series (details and approximation), introduced greater error and led to a drastic reduction in parsimony. Therefore, only the first three levels of decomposition were considered. The low frequency content of the signals (approximation) represents the identity of the signal. The high frequency content of the signals shows the signal details. Each of the details signals was trained with a single neural network, while each low frequency approximation signal was predicted using a SARIMA forecasting model.

An ANN consisting of a three layer feed forward perceptron structure and a back-propagation algorithm (BP) was determined to be appropriate for prediction of non-linearity and high frequency series such as details. An individual Multi Layer Perceptron (MLP) feed forward ANN was trained for each level's detail sub-series to predict normalized detail sub series (as output layer neuron (d_t)). To determine the optimal number of input neurons for the ANN model, d_t with lags 1 to 4 (d_{t-1}, d_{t-2}, $\ldots$, d_{t-4}) were considered for training each network. The neurons that increased model accuracy were selected as optimum input neurons and there was a single output neuron (normalized d_t). In order to obtain the parameters of the ANN

model (number of hidden nodes, the learning rate and the momentum value), the data set was divided into two parts: training and testing sets. A logistic sigmoid transfer function was applied in the hidden layer. The network was trained in 1000 epochs using the Levenberg–Marquardt learning algorithm with a learning rate of 0.001 and a momentum value of 0.7. In ANN modeling, the number of neurons in the input and output layers is determined by the number of input and output variables of the system under study, respectively. The model parameters were chosen by trial and error procedure and one hundred trials of random initial weights were considered to prevent the ANN method from falling in a local minimum. The optimal number of neurons in the hidden layer was obtained by trial and error by changing the number of neurons from 3 to 15. Each detail series was divided into three parts. The first 75% of the total series served as a training set, the second 10% served for cross validation and the final 15% was used for testing. The time series data were normalized to values between 0 and 1 prior to serving as an ANN input:

$$d_t = \frac{(d_i - d_{\min})}{(d_{\max} - d_{\min})} \qquad (8)$$

where d_t is the normalized detail sub series, $d_{\max}$ and $d_{\min}$ are the maximum and minimum of detail series, respectively, and d_i, is the detail in decomposition level i.

In the next step, the forecasting of approximate sub series in different decomposition levels, a SARIMA model was used. Finally, predicted details and approximate sub series were summed to forecast one-month-ahead precipitation. For example, in decomposition level 3, details sub series with the predicted

details ($D_{1(t)}^*$, $D_{2(t)}^*$ and $D_{3(t)}^*$) and predicted approximate sub series $A_{3(t)}^*$ were summed to obtain one month ahead precipitation.

$$R_{(t+1)} = A_3{}^*{}_{(t)} + D_3{}^*{}_{(t)} + D_2{}^*{}_{(t)} + D_1{}^*{}_{(t)} \qquad (9)$$

The coefficient of determination (R^2) and root mean square error ($RMSE$) were used to compare the performance of the different models:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(R_i - \hat{R}_l)^2}{\sum_{i=1}^{N}(\hat{R}_l - \bar{R})^2} \qquad (10)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(R_i - \hat{R}_l)}{N}} \qquad (11)$$

where R^2, N, R_i, $\hat{R}_l$, $\bar{R}$ are determination coefficient, number of observations, observed data, predicted values and mean of observed data, respectively.

STUDY AREA

The data used in this paper was obtained from a meteorological station situated on the Nahavand plain of western Iran's Hamedan province (34°12'N lat., 48°22'E long.). 644 km² in area, the Nahavand plain ranges in elevation between 1402 m and 3400 m above mean sea elevation. A 40-year (1970–2009) monthly precipitation time series was used in this study (Fig. 1). The monthly mean and maximum precipitation over this period were 44.4 mm and 261 mm, respectively. The mean monthly temperatures vary from –2.1°C in February to 23.67°C in July. A strong seasonality was clearly seen in the time series.
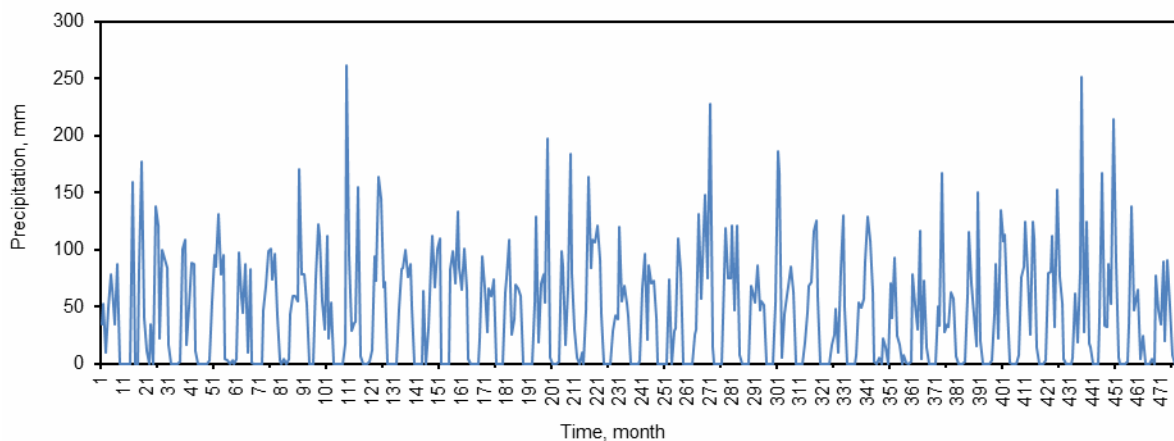

Fig. 1. Monthly precipitation time series; source: own study

RESULTS AND DISCUSSION

This research attempted to analyze the influences of the applied mother wavelet type as well as decomposition level on model performance. Hence, precipitation time series were broken up to one, two and three levels by three different types of wavelet trans-

forms, i.e., (1) Daubechies wavelet (db3), the most commonly used wavelet, (2) coif1 wavelet, which is characterized by three sharp peaks and (4) dmey wavelet, which has an irregular shape, in order to obtain detail and approximate sub series. For instance, the decomposition level 2 through the db$_3$ mother wavelet for a 100 month segment of the full series

(Fig. 2) showed the approximation signal to have preserved the overall form of the main signal, while the details signals showed most of the noise portion of the full time series.
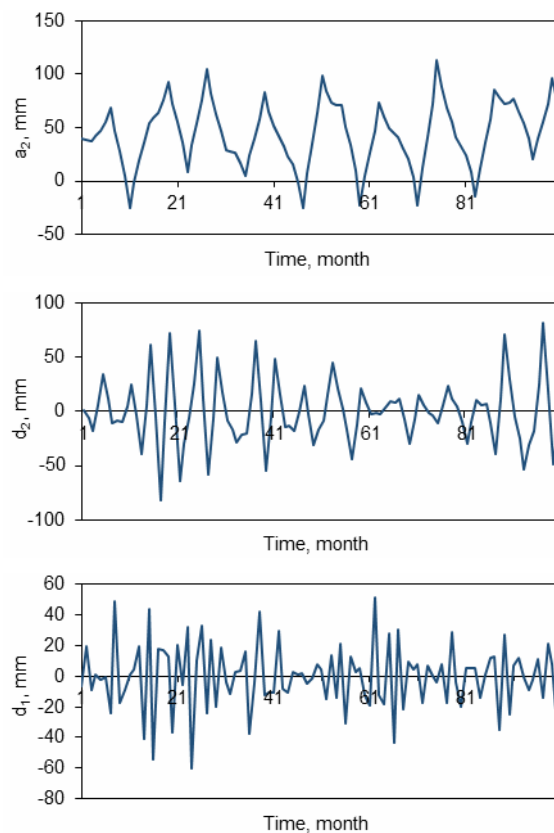

Fig. 2. Approximation and detail subseries of precipitation using the db3 wavelet (level 2); source: own study

Each of the details obtained through optimum lags were used to predict details over the testing period (Tab. 1). Prediction accuracy of details (d_1, d_2 and d_3) obtained from a db3 mother wavelet at decomposition level 3 exceeded that of other models in terms of evaluation criteria. Detail d_3 obtained from db3 with 4-19-1 ANN layers had the smallest error among ANN models. Nonetheless, results of prediction details d_1, d_2 and d_3 calculated in decomposition level 3 using the coif1 mother wavelet were similar to those obtained with db3 mother wavelet (Tab. 1).

In the next stage, a SARIMA (p, d, q) (P, D, Q) model was applied to each approximation component of precipitation. The parameters p, d, and q were estimated for each approximation. The ACF plot of the a_2 (level 2 approximation) sub-signal, resulting from a coif1 mother wavelet, showed the a_2 signal to be a seasonal wave with a period of 12 months (Fig. 3). Therefore, the SARIMA model with a 12 month seasonality and 2 sequence differencing was a potential model for the modeling of a_2. To select the exact selection of patterns for the SARIMA model's (p, d, q) (P, D, Q) values, various combinations were evaluat-

Table 1. Rainfall prediction accuracy, as assessed by R^2 and *RMSE*, for artificial neural networks (ANNs) varying in structure, along with the mother wavelet and wavelet decomposition level of the associated wavelet analysis

| Mother wavelet type	Decomposition level	Details	ANN structure	R^2 (Testing)	*RMSE* (Testing)
db3	1	d_1	1-5-1	0.74	1.75
db3	2	d_1	1-7-1	0.74	1.73
db3		d_2	2-16-1	0.80	1.65
db3	3	d_1	1-14-1	0.72	1.77
db3		d_2	2-8-1	0.85	1.01
db3		**d_3**	**4-19-1**	**0.88**	**0.99**
coif1	1	d_1	1-14-1	0.68	2.10
coif1	2	d_1	1-6-1	0.71	1.79
coif1		d_2	2-10-1	0.79	1.66
coif1	3	d_1	1-5-1	0.77	1.68
coif1		d_2	2-7-1	0.84	1.03
coif1		d_3	3-10-1	0.83	1.09
dmey	1	d_1	1-6-1	0.66	2.18
dmey	2	d_1	2-9-1	0.74	1.75
dmey		d_2	3-5-1	0.68	2.07
dmey	3	d_1	2-16-1	0.73	1.75
dmey		d_2	2-11-1	0.71	1.79
dmey		d_3	4-7-1	0.68	2.89

Source: own study.
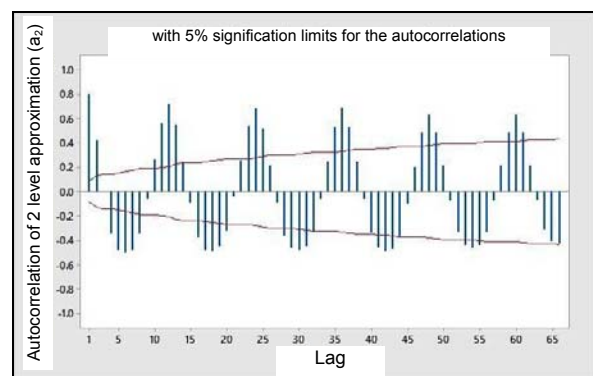
Fig. 3. Autocorrelation function (ACF) plot of approximation (a_2) using the coif1 wavelet; source: own study

ed based on the error criteria of mean absolute percentage error (*MAPE*), mean square error (*MSE*), and *RMSE*. A SARIMA (3, 0, 0) (3, 2, 3)$_{12}$ model showed the lowest error criteria. The coefficients of the SARIMA (3, 0, 0) (3, 2, 3)$_{12}$ model were obtained using Minitab software. Considering this model's residuals values, the bell shape of its frequency histogram, the appropriate adaptation of observed and theoretical values as well the ACF and PACF (partial autocorrelation function) being inside the independent boundary (Fig. 4), confirmed that the model was well fitted.
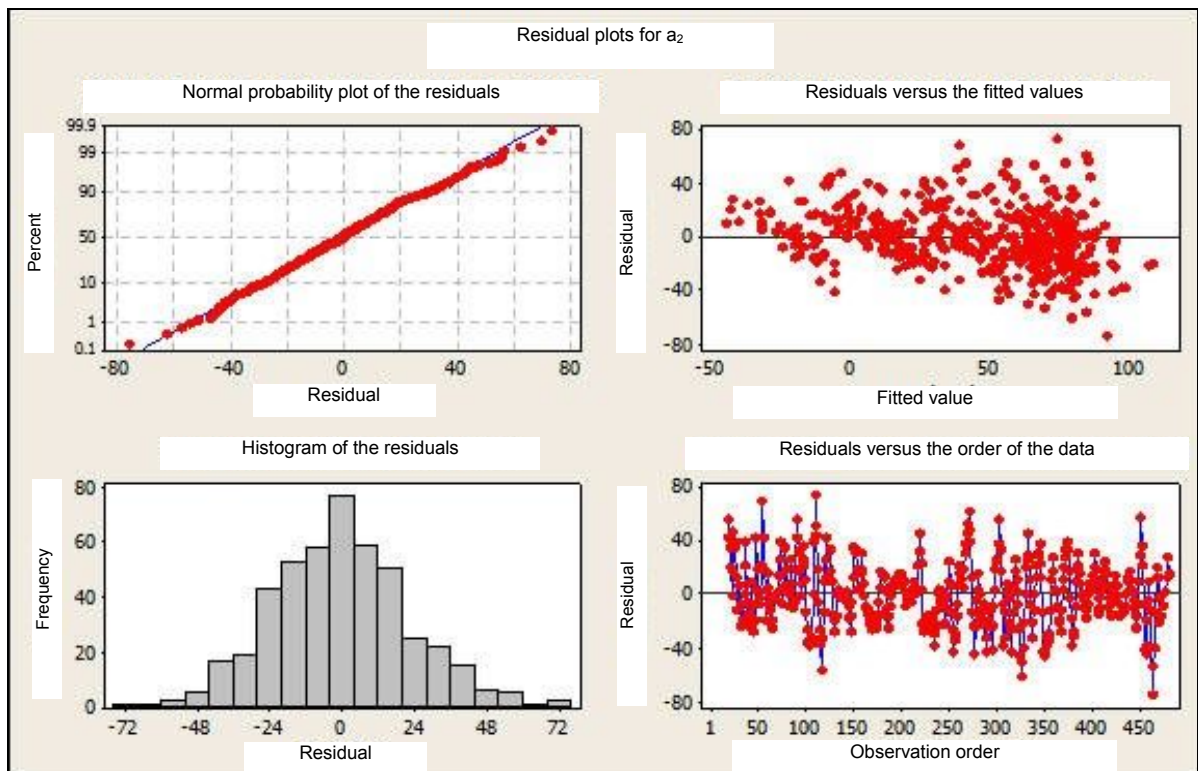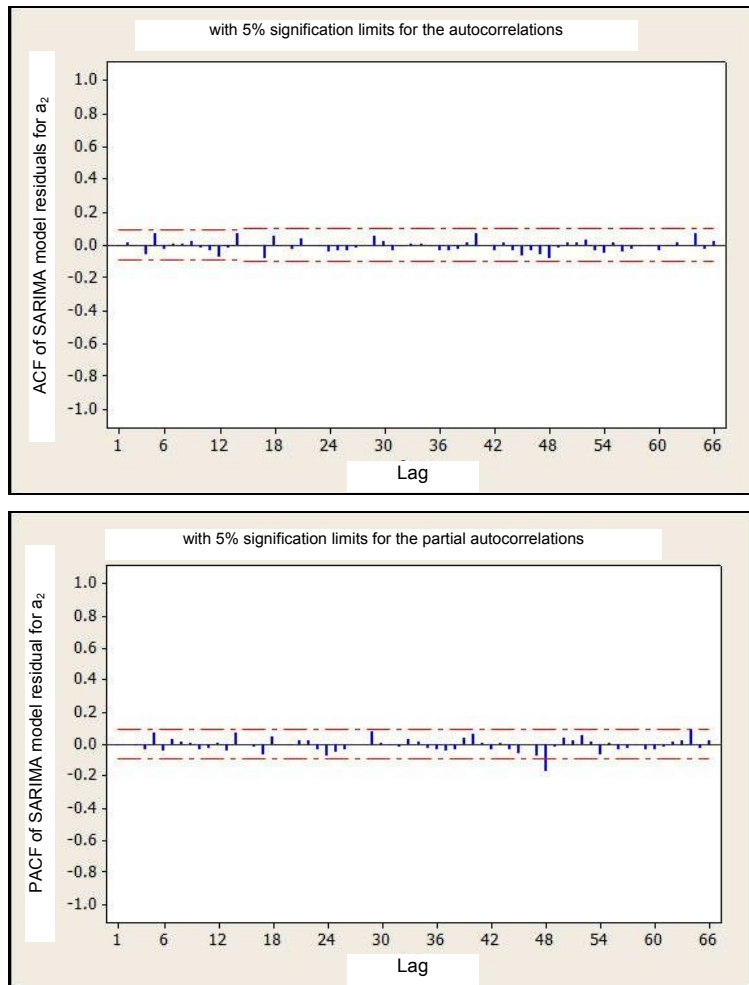
Fig. 4. Normal probability and histogram of residual; autocorrelation function (ACF)
and partial autocorrelation function (PACF) plot for residual of SARIMA model in a_2; source: own study

Similarly, appropriate SARIMA models were found for a_1, a_2, a_3, obtained through various mother wavelets (coif1, db3, dmey). In the next step the prediction of approximate sub series by the SARIMA model was tested using the last 15% of data (72 months). The accuracy criterion of R^2 for the SARIMA model (Tab. 2) showed that an a_2 approximate sub series with structure SARIMA $(3,0,0)(3,2,3)_{12}$ obtained from a coif1 mother wavelet showed the greatest accuracy amongst the SARIMA models ($R^2 = 0.98$), while an a_3 with a SARIMA $(3,0,0)$ $(3,1,3)_{12}$ structure, developed from a db3 mother wavelet, was the second most accurate ($R^2 = 0.97$).

Table 2. Prediction accuracy of seasonal, auto, regressive integrating average (SARIMA) models for approximation

Approximation	Mother wavelet type	SARIMA	R^2
a_1	coif1	$(3,0,0)(2,1,2)_{12}$	0.85
a_2	**coif1**	**$(3,0,0)(3,2,3)_{12}$**	**0.98**
a_3	coif1	$(3,1,0)(3,2,3)_{12}$	0.96
a_1	db3	$(3,0,0)(2,2,1)_{12}$	0.86
a_2	db3	$(3,0,0)(3,2,3)_{12}$	0.95
a_3	db3	$(3,0,0)(3,1,3)_{12}$	0.97
a_1	dmey	$(3,0,0)(3,2,3)_{12}$	0.82
a_2	dmey	$(3,0,0)(3,2,3)_{12}$	0.73
a_3	dmey	$(2,0,0)(2,1,3)_{12}$	0.96

Source: own study.

To obtain the predicted precipitation, reconstruction was carried out by summing the predicted details and approximation. Table 3 illustrates the adaptation of the SARIMA model and ANN structure for various mother wavelets. Table 3 indicates that the models obtained from decomposition level 3 through db3 performed better than other models in terms of R^2 and *RMSE*, while models obtained by level 2 and 3 decomposition using coif1 showed a similar accuracy as models obtained from db3 (Tab. 3).

Table 3. Accuracy assessment of precipitation models obtained using a wavelet-SARIMA-ANN (WSA) model

Mother wavelet type	Decomposition level	R^2	*RMSE*
db3	1	0.650	2.19
db3	2	0.950	1.23
db3	3	0.970	1.19
coif1	1	0.680	2.34
coif1	2	0.965	1.22
coif1	3	0.967	1.22
dmey	1	0.630	2.22
dmey	2	0.860	1.45
dmey	3	0.810	1.78

Source: own study.

To further evaluate the proposed wavelet--SARIMA-ANN (WSA) models' accuracy, plots of predicted time series were plotted against observed values at decomposition levels of 2 and 3 for mother wavelets db3, dmey and coif1 (Fig. 5, 6 and 7, respectively). Data computed from decomposition level 3
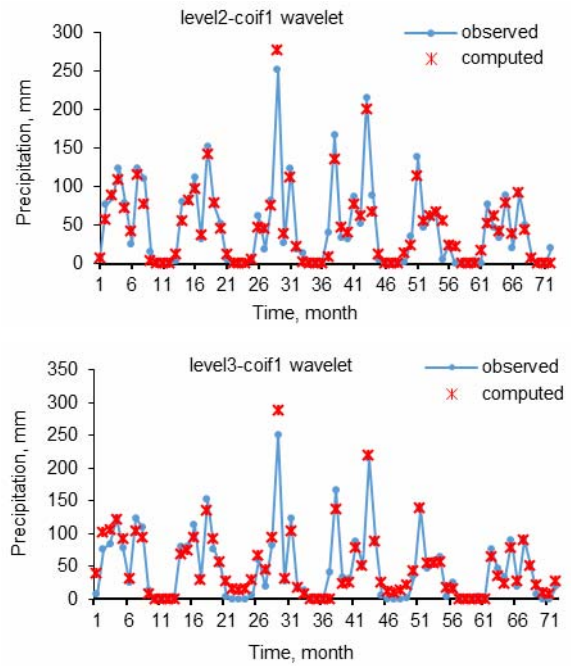

Fig. 5. Plot of observed versus computed precipitation using wavelets of level 2 and 3 using the coif1 wavelet; source: own study
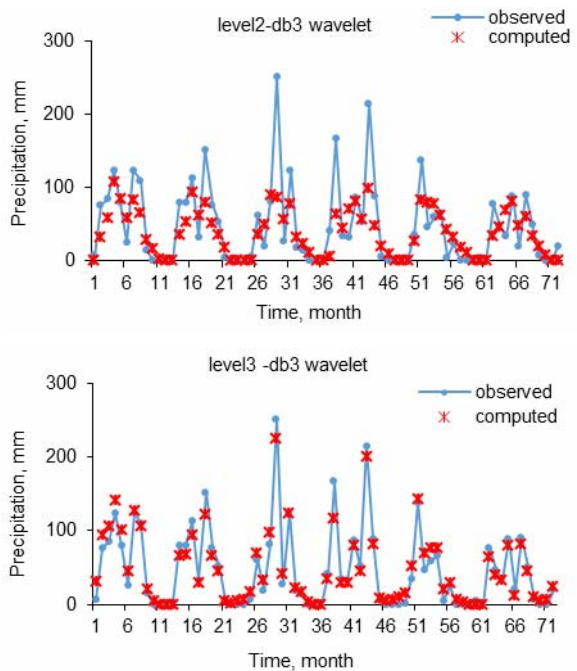

Fig. 6. Plot of observed versus computed precipitation using wavelet transform and levels 2 and 3 using the db3 wavelet; source: own study

through db3 showed the best correspondence between modelled and observed data overall, particularly for peak points, which are important in water resource management. While models computed at decomposition levels 2 and 3 through coif1 also showed a close overlap with observed data (Fig. 5), models computed at level 3 through dmey were poor compared to other models, particularly in terms of forecasting peak points.
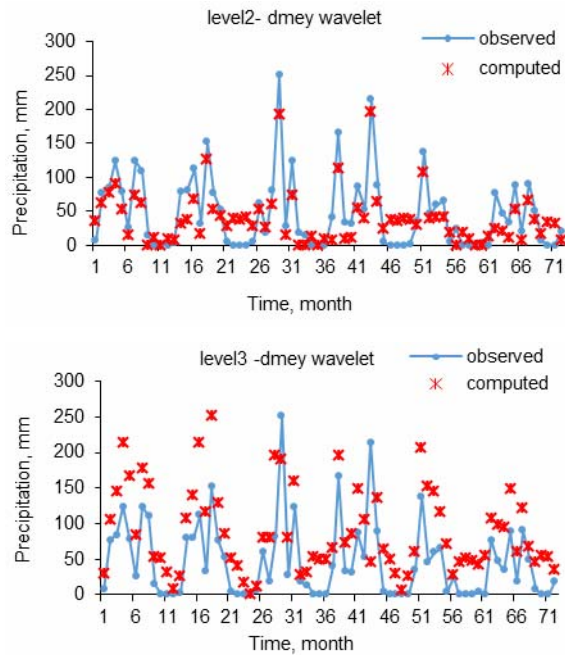
Fig. 7. Plot of observed versus computed precipitation using wavelet transform at levels 2 and 3 using the dmey wavelet; source: own study

In order to evaluate the predictive ability of the proposed WSA model, some comparisons for modeling of the precipitation process were done with SARIMA and W-ANN models alone (Tab. 4). The accuracy of the W-ANN model (details and approximation are imposed to ANN) using mother wavelets (db3, coif1, dmey) at all decomposition levels was less accurate than that of the WSA (Tab. 4). Moreover, the SARIMA $(3, 0, 3) (0, 1, 2)_{12}$ model, with an R^2 of 0.69, exhibited poor precipitation modelling ability.

Table 4. Results of prediction using wavelet-artificial neural network (W-ANN) model

Wavelet type	Decomposition level	R^2	*RMSE*
db3	1	0.68	2.35
db3	2	0.78	2.01
db3	3	0.70	2.33
coif1	1	0.77	2.03
coif1	2	0.89	1.43
coif1	3	0.77	2.04
dmey	1	0.64	2.50
dmey	2	0.80	1.99
dmey	3	0.77	2.03

Source: own study.

CONCLUSION

An attempt was made to model monthly precipitation using the properties of noisy wave simplification through wavelet decomposition procedures. Overall, the extracted cycles from the hydrological wave obtained through the wavelet transform represented physical concepts and particular climatic char-

acteristics. Short and long term features of hydrological events could be identified through analysis of these frequencies (low and high frequencies). The monthly precipitation at the Nahavand station (Hamedan province, Iran) was considered for this purpose. Precipitation data were decomposed to 1, 2 and 3 levels. Every level contained an approximation representing the smooth view of the original wave and was compatible with modelling using SARIMA group models. In contrast, the details which were of a noisy nature could be best analyzed by artificial intelligence procedures (i.e. ANNs). Therefore each detail was trained with an ANN, while approximation prediction was performed with a SARIMA model. Predicting approximations with the SARIMA model and details with an ANN increased accuracy (greater R^2 between observed and computed) of the WSA model compared with SARIMA or W-ANN models alone. Decomposition with db3 and coif1 wavelets led to more accurate predictions than decomposition with the demey wavelet. This may be attributable to a similarity between the precipitation time series and the coif and db3 mother wavelets involved in the effective separation of low-frequency and high frequency events in the precipitation time series. This led to the greater accuracy of the WSA model compared to the W-ANN models. The WSA model also showed a greater precision of results in terms of precipitation peaks, an important element in precipitation modeling. In investigating the effect of decomposition level on model performance, it was found that going from 2 to 3 decomposition levels raised the correlation between observed and estimated data, but no significant difference was found between predictions from 2 and 3 level models. Moreover, an increase in the waves (details and approximation) beyond three induced a greater error and a significant reduction in parsimony, and therefore generated no benefits. As mentioned earlier, precipitation is an important variable in the context of climate change, and as such its prediction is important in facilitating the transition to more integrated and adaptive water resources planning and management [BUTLER *et al.* 2015; HALBE *et al.* 2014; INAM *et al.* 2015; KOLINJIVADI *et al.* 2014; STRAITH *et al.* 2014].

REFERENCES

ADAMOWSKI J., ADAMOWSKI K., BOUGADIS J. 2010. Influence of trend on short duration design storms. Water Resources Management. Vol. 24 p. 401–413.

ADAMOWSKI J., CHAN H.F.A. 2011. Wavelet neural network conjunction model for groundwater level forecasting. Journal of Hydrology. Vol. 407. Iss. 1–4 p. 28–40. DOI: 10.1016/j.jhydrol.2011.06.013

ADAMOWSKI J., CHAN H., PRASHER S., SHARDA V.N. 2012a. Comparison of multivariate adaptive regression splines

with coupled wavelet transform artificial neural networks for runoff forecasting in Himalayan micro-watersheds with limited data. Journal of Hydroinformatics. Vol. 3 p. 731–744.

ADAMOWSKI J., PROKOPH A., ADAMOWSKI K. 2012b. Influence of the 11 year solar cycle on annual streamflow maxima in Southern Canada. Journal of Hydrology. Vol. 442 p. 55–62.

ADAMOWSKI J., PROKOPH A., ADAMOWSKI K. 2013. A spectral analysis based methodology to detect climatological influences on daily urban water demand. Mathematical Geosciences. Vol. 45 p. 49–68.

ADAMOWSKI J., SUN K. 2010. Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. Journal of Hydrology. Vol. 390. Iss. 1–2 p. 85–91. DOI: 10.1016/j.jhydrol.2010.06.033

ADAMOWSKI K., PROKOPH A., ADAMOWSKI J. 2009. Development of a new method of wavelet aided trend detection and estimation. Hydrological Processes. Vol. 23 p. 2686–2696.

AKANSU A.N., SERDIJN W.A., SELESNICK I.W. 2010. Emerging applications of wavelets: A review. Physical Communications. Vol. 3. Iss. 1 p. 1–18. DOI: 10.1016/j.phycom.2009.07.001.

ARAGHI A., ADAMOWSKI J., NALLEY D., MALARD J. 2015. Using wavelet transforms to estimate surface temperature trends and dominant periodicities in Iran based on gridded reanalysis data. Atmospheric Research. Vol. 155 p. 52–72.

BELAYNEH A., ADAMOWSKI J., KHALIL B., OZGA-ZIELINSKI B. 2014. Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet-support vector regression models. Journal of Hydrology. Vol. 508 p. 418–429.

BOX G.E.P., JENKINS G.M. 1976. Time series analysis: Forecasting and control. 4th ed. San Francisco, CA. Holden Day.

BUTLER C., ADAMOWSKI J. 2015. Empowering marginalized communities in water resources management: Addressing inequitable practices in Participatory Model Building. Journal of Environmental Management. Vol. 153 p. 153–162.

CAMPISI S., ADAMOWSKI J., ORON G. 2012. Forecasting urban water demand via wavelet- denoising and neural network models. Case study: City of Syracuse, Italy. Water Resources Management. Vol. 26 p. 3539–3558.

CANNAS B., FANNI A., SEE L., SIAS G. 2006. Data preprocessing for river flow forecasting using neural networks: wavelet transforms and data partitioning. Physics and Chemistry of the Earth. Parts A/B/C. Vol. 31. Iss. 18 p. 1164–1171. DOI:10.1016/j.pce.2006.03.020.

GOYAL M., BHARTI B., QUILTY J., ADAMOWSKI J., PANDEY A. 2014 Modeling of daily pan evaporation in sub tropical climates using ANN, LS-SVR, fuzzy logic, and ANFIS. Expert Systems with Applications, 41, 11 p. 5267–5276.

HAIDARY A., AMIRI B.J., ADAMOWSKI J., FOHRER N., NAKANE K. 2013. Assessing the impacts of four land use types on the water quality of wetlands in Japan. Water Resources Management. Vol. 27 p. 2217–2229.

HALBE J., ADAMOWSKI J., BENNETT E., PAHL-WOSTL C., FARAHBAKHSH K. 2014. Functional organization analysis for the design of sustainable engineering systems. Ecological Engineering. Vol. 73 p. 80–91.

HALBE J., PAHL-WOSTL C., SENDZIMIR J., ADAMOWSKI J. 2013. Towards adaptive and integrated management

paradigms to meet the challenges of water governance. Water Science and Technology: Water Supply. Vol. 67. Iss. 11 p. 2651–2660.

INAM A., ADAMOWSKI J., HALBE J., PRASHER S. 2015. Using causal loop diagrams for the initialization of stakeholder engagement in soil salinity management in agricultural watersheds in developing countries: A case study in the Rechna Doab watershed, Pakistan. Journal of Environmental Management. Vol. 152 p. 251–267.

KARRAN D., MORIN E., ADAMOWSKI J. 2014. Multi-step streamflow forecasting using data-driven non-linear methods in contrasting climate regimes. Journal of Hydroinformatics. Vol. 16. Iss. 3 p. 671–689.

KIM T.W., VALDES J.B. 2003. Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. Journal of Hydrologic Engineering. Vol. 8. Iss. 6 p. 319–328. DOI: 10.1061/(ASCE)1084-0699(2003)8:6(319).

KISI O., CIMEN M. 2012. Precipitation forecasting by using wavelet-support vector machine conjunction model. Engineering Applications of Artificial Intelligence. Vol. 25. Iss. 4 p. 783–792. DOI: 10.1016/j.engappai.2011.11.003.

KISI O., SHIRI J. 2011. Precipitation forecasting using wavelet-genetic programming and wavelet-neuro-fuzzy conjunction models. Water Resource Management. Vol. 25. Iss. 13 p. 3135–3152. DOI: 10.1007/s11269-011-9849-3.

KOLINJIVADI V., ADAMOWSKI J., KOSOY N. 2014. Recasting payments for ecosystem services (PES) in water resource management: A novel institutional approach. Ecosystem Services. Vol. 10 p. 144–154.

LU R.Y. 2002. Decomposition of interdecadal and interannual components for North China rain fall in rainy season. Chinese Journal of Atmosphere. Vol. 26 p. 611–624 (in Chinese).

NALLEY D., ADAMOWSKI J., KHALIL B. 2012. Using discrete wavelet transforms to analyze trends in streamflow and precipitation in Quebec and Ontario (1954–2008). Journal of Hydrology. Vol. 475 p. 204–228.

NALLEY D., ADAMOWSKI J., KHALIL B., OZGA-ZIELINSKI B. 2013. Trend detection in surface air temperature in Ontario and Quebec, Canada during 1967–2006 using the discrete wavelet transform. Atmospheric Research. Vol. 132/133 p. 375–398.

NASON G.P., VON SACHS R. 1999. Wavelets in time series analysis. Philosophical Transactions of the Royal Society A. Vol. 357. Iss. 1760 p. 2511–2526. DOI: 10.1098/rsta.1999.0445

NOURANI V., BAGHANAM A., ADAMOWSKI J., KISI O. 2014. Applications of hybrid wavelet-artificial intelligence models in hydrology: A review. Journal of Hydrology. Vol. 514 p. 358–377.

PINGALE S., KHARE D., JAT M., ADAMOWSKI J. 2014. Spatial and temporal trends of mean and extreme rainfall and temperature for the 33 urban centres of the arid and semi-arid state of Rajasthan, India. Atmospheric Research. Vol. 138 p. 73–90.

RATHINASAMY M., ADAMOWSKI J., KHOSA R. 2014. Multiscale streamflow forecasting using a new Bayesian model average based ensemble multi-wavelet Volterra nonlinear method. Journal of Hydrology. Vol. 507 p. 186–200.

RATHINASAMY M., KHOSA R., ADAMOWSKI J., CH S., PARTHEEPAN G., ANAND J., NARSIMLU B. 2015. Wavelet-based multiscale performance analysis: An approach to assess and improve hydrological models. Water Re-

sources Research. Vol. 50. Iss. 12 p. 9721–9737. DOI: 10.1002/2013WR014650.

SAADAT H., ADAMOWSKI J., BONNELL R., SHARIFI F., NAMDAR M., ALE-EBRAHIM S. 2011. Land use and land cover classification over a large area in Iran based on single date analysis of satellite imagery. Journal of Photogrammetry and Remote Sensing. Vol. 66 p. 608–619.

SALAS J.D., DELLEUR J.W., YEVJEVICH V., LANE W.L. 1980. Applied modeling of hydrological time series. Littleton, CO. Water Resources Publications. ISBN 978-09-18334-37-4 pp. 484.

SANTOS C.A.G., MORAIS B.S., SILVA G.B.L. 2009. Drought forecast using an artificial neural network for three hydrological zones in San Francisco River basin, Brazil. IAHS Publication. Vol. 333 p. 302–312.

SANTOS C.A.G., SRINIVASAN V.S., SUZUKI K., WATANABE M. 2003. Application of an optimization technique to a physically based erosion model. Hydrological Processes. Vol. 17. Iss. 5 p. 989–1003. DOI: 10.1002/hyp.1176.

SELESNICK I.W., BARANIUK R.G., KINGSBURY N.C. 2005. The dual-tree complex wavelet transform. Signal Processing Magazine, IEEE. Vol. 22. Iss. 6 p. 123–151. DOI: 10.1109/MSP.2005.1550194

STRAITH D., ADAMOWSKI J., REILLY K. 2014. Exploring the attributes, strategies and contextual knowledge of champions of change in the Canadian water sector. Canadian Water Resources Journal. Vol. 39. Iss. 3 p. 255–269.

TANTANEE S., PATAMATAMMAKUL S., OKI T., SRIBOONLUE V., PREMPREE T. 2005. Coupled wavelet-autoregressive model for annual rainfall prediction. Journal of Environmental Hydrology. Vol. 13. Paper 18 p. 1–8.

TIWARI M., ADAMOWSKI J. 2014. Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. Water Resources Research. Vol. 49. Iss. 10 p. 6486–6507.

TIWARI M., ADAMOWSKI J. 2015. Medium-term urban water demand forecasting with limited data using an ensemble wavelet-bootstrap machine-learning approach. Journal of Water Resources Planning and Management. Vol. 141. No. 2.

Maryam SHAFAEI, Jan ADAMOWSKI Ahmad FAKHERI-FARD, Yagob DINPASHOH, Kazimierz ADAMOWSKI

Hybrydowy model wavelet-SARIMA-ANN do prognozowania opadów

STRESZCZENIE

Słowa kluczowe: *gospodarka zasobami wodnymi, metoda wavelet, prognozowanie opadów, sezonowa zintegrowana autoregresja z ruchomą średnią, sztuczne sieci neuronowe*

Prognozowanie opadów, ze względu na ich znaczenie w gospodarce zasobami wodnymi, szczególnie w zmniejszaniu ryzyka powodzi czy susz, było już przedmiotem wielu badań. Serie miesięcznych opadów mają właściwości nieliniowe i różne skale czasowe, w związku z czym przetestowano różne metody: wavelet, metodę zintegrowanej sezonowej autoregresji z ruchomą średnią (SARIMA) i hybrydową metodę sztucznych sieci neuronowych (ANN) pod kątem ich zdolności do dokładnego przewidywania miesięcznych opadów. Czterdziestoletnią (1970–2009) serię opadów z irańskiej stacji meteorologicznej w Nahavand (34°12'N, 48°22'E) rozłożono na jedną podserię o niskiej częstotliwości i kilka podserii o wysokiej częstotliwości występowania opadów przez transformację falkową. Podserie o niskiej częstotliwości prognozowano za pomocą modelu SARIMA, podczas gdy podserie o wysokiej częstotliwości prognozowano, stosując ANN. Na koniec prognozowane podserie zrekonstruowano celem przewidywania opadów w poszczególnych miesiącach w przyszłości. Porównanie wartości generowanych przez model z danymi z obserwacji wykazało lepszą dokładność prognozowania opadów za pomocą modelu wavelet-SARIMA-ANN niż za pomocą modeli wavelet-ANN i wavelet-SARIMA.