Environ Monit Assess (2021) 193:355
https://doi.org/10.1007/s10661-021-09060-8

Multivariate modeling of pan evaporation in monthly temporal resolution using a hybrid evolutionary data-driven method (case study: Urmia Lake and Gavkhouni basins)

Alireza Emadi · Sarvin Zamanzad-Ghavidel · Sina Fazeli · Soheila Zarei · Ali Rashid-Niaghi

Received: 12 February 2021 / Accepted: 11 April 2021

Abstract Evaporation is an important meteorological variable that has a great impact on water resources. In the current research, climatology data, and seasonal coefficient have been used to estimate monthly pan evaporation (E_{pan}) for 2005–2018 study years at four selected stations of the Urmia Lake basin with Dsa and six selected stations of Gavkhouni basin with Bsk climate categories, in Iran. Estimation of monthly E_{pan} was performed using data-driven methods such as artificial neural networks (ANNs), adaptive neuro-fuzzy inference system (ANFIS), and gene expression programming (GEP) as well as wavelet-hybrids (WANN, WANFIS, and WGEP). Based on the evaluation criteria, the WGEP model performance was better than the other models in estimating the monthly E_{pan}. The results indicated that WGEP and ANN are the best and poorest models for all stations without affecting the climate condition of basins. The values of RMSE for WGEP model for stations of Urmia Lake and Gavkhouni basins were varied from 15.839 to 26.727 and 20.651 to 70.318, respectively. Also, the values of RMSE for ANN model for stations of Urmia Lake and Gavkhouni basins were varied from 29.397 to 38.452 and 30.635 to 85.237, respectively. The model's performance was improved as a result of considering the data noise elimination and applying seasonal coefficient to estimate E_{pan} of various climatic conditions. This study with presenting mathematical equations for estimating monthly E_{pan} has a significant impact on the management and planning of water resources policymakers in the future.

Keywords Climate condition · Data-driven methods · Pan evaporation · Simulation · Wavelet theory

A. Emadi (✉)
Department of Water Engineering, Sari Agricultural Science And Natural Resources University, Sari, Iran
e-mail: emadia355@yahoo.com

S. Zamanzad-Ghavidel
Department of Irrigation & Reclamation Engineering, Faculty of Agriculture Engineering & Technology, College of Agriculture & Natural Resources, University of Tehran, Iran's National Science Foundation (INSF), Tehran, Iran
e-mail: snzghavidel@ut.ac.ir

S. Fazeli · S. Zarei
Department of Irrigation & Reclamation Engineering, Faculty of Agriculture Engineering & Technology, College of Agriculture & Natural Resources, University of Tehran, Tehran, Iran
e-mail: s.fazeli@ut.ac.ir

S. Zarei
e-mail: soheila_zareie@yahoo.com

A. Rashid-Niaghi
Benson Hill, St. Louis, MO, USA
e-mail: niaghi@umn.edu

Introduction

Evaporation is one of the main components of the water process and climate changes in which an accurate estimation of evaporation can enhance water resource

Published online: 24 May 2021

management and improve the monitoring network. Evaporation happens as a result of available energy and a vapor pressure deficit between a water surface and the overlying atmosphere (Moghaddamnia et al., 2009). Many climatic parameters influence the rate of evaporation, such as air temperature, precipitation, relative humidity, solar radiation, and wind speed (Azamathulla et al., 2018; Perera et al., 2020a, b).

There are direct and indirect approaches available for calculating and forecasting evaporation. Using pan evaporation (E_{pan}) (Class A pan, US Weather Bureau) is one of the most important direct methods for evaporation assessment (Eslamian et al., 2008). Modeling E_{pan} plays an important role in water budgeting for various weather conditions (Khaniya et al., 2020; Majidi et al., 2015; Sanikhani et al., 2012). Indirect methods such as water and energy budget and Penman–Monteith are based on different meteorological variables (Ghaemi et al., 2019; McColl, 2020; Or & Lehmann, 2019). However, the amount of required data and difficulty in estimation of the unknown meteorological variables discouraged researchers from using indirect methods (Burt et al., 2005; Khatibi et al., 2020; Sanikhani et al., 2015). Evaporation is a nonlinear process, so it is difficult to elicit a valuable mathematical equation to address all the physical processes (Malik et al., 2017; Tezel & Buyukyildiz, 2016; Trajkovic & Kolakovic, 2010). Therefore, the application of computer models and algorithms that can overcome the prior knowledge requirement on physical processes as well as the required amount of data is recommended (Tan et al., 2018; Tezel & Buyukyildiz, 2016; Wang et al., 2017).

Recent efforts of using artificial intelligence (AI) models have shown great progress in E_{pan} estimation (Tao et al., 2018). Several approaches of AI model that applied to estimate E_{pan} are artificial neural network (ANN), co-active neuro-fuzzy inference system (CANFIS), support vector machine (SVM), gene expression programming (GEP), and complementary wavelet-AI (Adnan et al., 2019; Guven & Kisi, 2013; Kisi & Heddam, 2019; Qasem et al., 2019; Qutbudin et al., 2019; Sebbar et al., 2019; Yaseen et al., 2015). The ANFIS, ANN, and Stephens-Stewart (SS) method have been applied by Kisi (2006) for daily E_{pan} modeling at Arcata-Euraka (moist and cool climate, CA, USA) and Daggett station (medium to the high desert climate, CA, USA) and reported the best performance of ANFIS model in comparison with the ANN and SS models. Also, the ANN model indicated a better performance compared to the SS method. Evaporation was simulated by Moghaddamnia et al. (2009) using the ANN, ANFIS, and empirical models while applying the gamma-test (GT) to select suitable input variables for the arid region. They found that the ANFIS and ANN models have better performance than the empirical formulas on evaporation estimation. Results of ANN, multiple linear regressions (MLR), Penman, Priestley-Taylor, and SS models have been compared for estimating the daily E_{pan} for semiarid and sub-tropical regions with hot dry summer and cold winter in which the ANN model showed an acceptable performance among other models (Shirsath & Singh, 2010). The ANN, CANFIS, and MLR have been used to simulate E_{pan} in a region with a sub-humid and sub-tropical climate (Malik & Kumar, 2015). Results included that the performance of the ANN model was superior to the CANFIS and MLR models. In a study by Keshtegar et al. (2016) for daily E_{pan} simulation, general regression analysis showed the best performance for ANFIS and M5Tree models for a region with a very hot and dry climate. The capability of ANFIS and ANN models for E_{pan} estimation was performed by Shiri (2019) for various regions in the USA and at the end; these models with ancillary data were recommended when there was no ability to achieve the local meteorological data.

The ability of AI-based models (neural network autoregressive with exogenous input (NNARX), GEP, and ANFIS) has been investigated by Zounemat-Kermani et al. (2019) for two stations, namely Gaziantep and Adiyaman, with different geographical locations but dry and semi-arid climate. Results of the comparison showed that the ANFIS model had better results in Gaziantep Station, while the NNARX model performed better in estimating E_{pan} values for the Adiyaman Station. Multiple-model ANN (MM-ANN), multivariate adaptive regression spline (MARS), SVM, multi-gene genetic programming (MGGP), and M5Tree were compared to simulate the E_{pan} on a monthly scale (Malik et al., 2020). The results included that MM-ANN-1 and MGGP-1 were more successful than the other techniques. The hybrid model combines two or more models and coupling is a different technique to make a hybrid model (Chia et al., 2020). This way has recently received more attention in the hydrology and climate process since the particular features of each technique may be able to better capture particular patterns in the data series (Ghaemi et al., 2019). Pammar and Deka (2017) attempted to explore hybrid modeling using discrete wavelet transform (DWT) and SVM for E_{pan} estimation without data noise elimination at two stations in India which represent different climatic zones. Also, they used the GT and grid search to select the best input–output combinations for the models. The E_{pan} estimates were very promising regarding expected accuracy in the associated fields. Kisi and Alizamir (2018) investigated the applicability of the wavelet extreme learning machine (WELM) model which uses discrete wavelet transform and ELM methods in estimating daily reference evapotranspiration (ET_0) at two stations, Ankara and Kirikkale, located in the central Anatolia region of Turkey. The results showed that the wavelet conjunction models (e.g., WELM and WANN) have better accuracy compared to the single models and the WELM model was found to be the best model in estimating ET_0. Rezaie-Balf et al. (2019) proposed two hybrid models, ensemble empirical mode decomposition (EEMD) coupled with SVM (EEMD-SVM) and EEMD-model tree (EEMD-MT) to improve the performance of monthly E_{pan} estimation without data noise elimination at two stations with a dry and semi-arid climate at Turkey. Results showed that newly proposed models gain higher accuracy compared to the conventional approaches. Jing et al. (2019) reviewed data-driven algorithms to estimate the ET_0. Their literature reviews showed the ability of providing mathematical equations and simple application of the GP and GEP models in ET_0 estimation for various range of climate. Chia et al. (2020) collected the results of studies on the estimation of ET_0 between 2011 and 2019. They declared that based on the results, hybrid models are used to simulate hydrological processes, such as ET_0, to overcome the errors of individual models. Their results indicate that the AI approach is the best possible solution to show the relationship between climatic parameters and ET_0. Yaseen et al. (2020) used machine learning (ML) models in two different stations in arid and semi-arid regions in Mosul and Baghdad, Iraq for estimating evaporation, comparing the applied model performance, and evaluating the role of climatic factors. Adnan et al. (2021) used several models such as GMDHNN, MARS, and M5Tree to simulate monthly ET_0 at two stations in Mediterranean region in Turkey. They compared the results of mentioned developed models with the results of empirical formulations including Hargreaves-Samani (HS), Calibrated-HS (CHS), and Stephens-Stewart (SS) models. The results of this study showed that GMDHNN and CHS have better accuracy and worst estimates, respectively.

Despite the great improvement of AI-based model and hydrological parameter estimation, most of the previous studies focused on the singularity of the models or assessing the elementary level of the hybrid model without data noise elimination and seasonal coefficient (SC) evaluation. Besides, the number of studied stations, climate zones, and applied models were limited. Therefore, this study was conducted to assess three single AI-based models (ANN, ANFIS, and GEP) as well as their wavelet-hybrid models (WANN, WANFIS, and WGEP) for E_{pan} estimation by considering two different climate zones with various stations. Considering the innovation of data noise elimination and applying seasonal coefficient to estimate E_{pan} of various climatic conditions make current research distinguished compared to the previous literature. Considering different climate zones makes a better comparison of the performance of the models.

Methodology

Case study and dataset

The Urmia Lake and Gavkhouni basins, with an area of 51.876 and 41.552 km^2, are located in the Northwest and central part of Iran, respectively, and were selected to study monthly E_{pan} (Fig. 1). According to the Köppen-Geiger climate classification method, the Urmia Lake has snow dry and hot summer (Dsa), and the Gavkhouni basin has arid desert cold (Bsk) climate type. Zarrinehrood, Siminehrood, and Aji Chai rivers are the main symbols of water resources of Urmia Lake. Zayandehrud River, as the largest central river in Iran, is one of the prominent symbols of the Gavkhouni basin. Both basins have a great impact on industrial, agricultural, and livestock activities of the northwestern and central part of Iran.

To estimate monthly E_{pan}, monthly precipitation (P), average temperature (T_{mean}), minimum temperature (T_{min}), maximum temperatures (T_{max}), relative humidity at noon (H_{12}) and 6 p.m. (H_{18}), and wind speed (U) data were collected from the Iranian Water

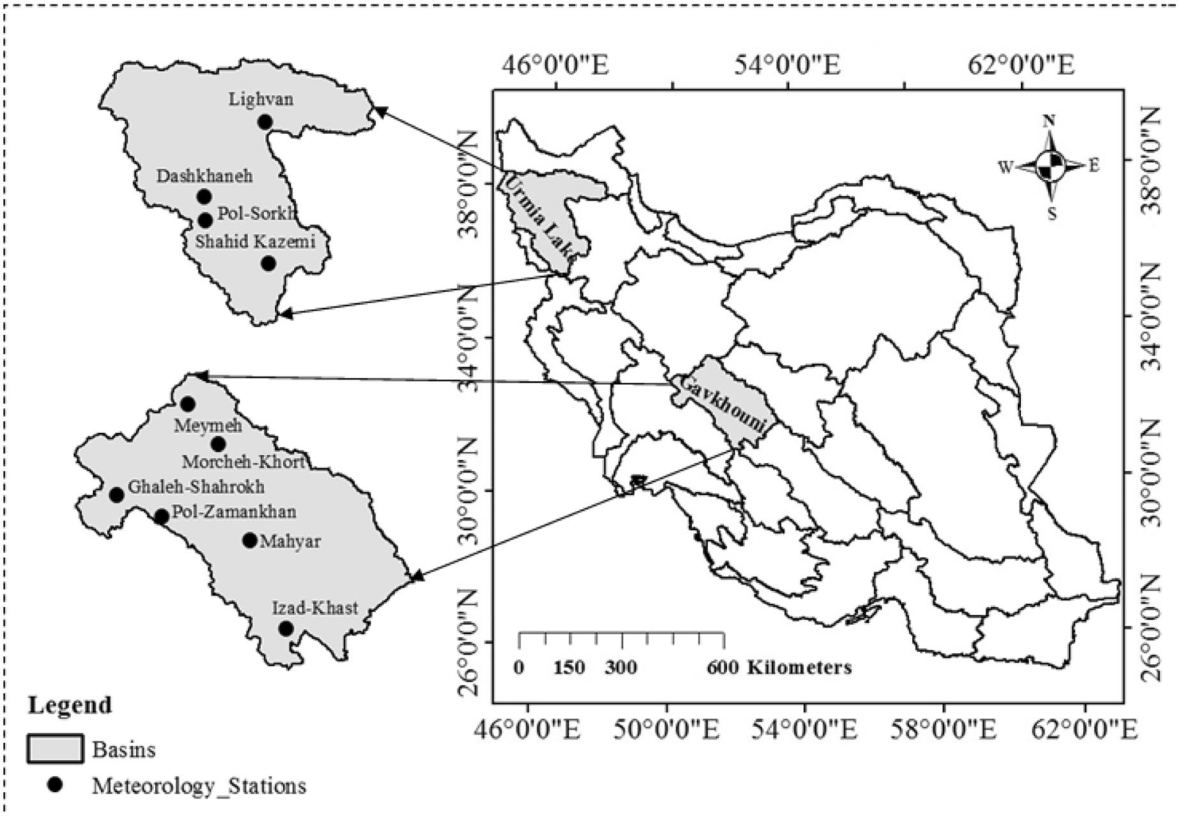


Fig. 1 The geographical location of the study basins including synoptic stations

Resources Management Company (http://wrbs.wrm.ir/). The collected data covered 2005–2018 study years including four selected stations of the Urmia Lake basin, namely Shahid Kazemi, Pol-Sorkh, Dashkhaneh, and Lighvan. A similar window of data was achieved for the Gavkhouni basin that included six selected stations, namely Ghaleh-Shahrokh, Mahyar, Izad-Khast, Morcheh-Khort, Meymeh, and Pol-Zamankhan. The location of the studied basins and selected meteorological stations is shown in Fig. 1.

To analyze the suitability of the selected basin in terms of climate conditions, the distribution of stations was assessed at a 99% probability level using F-computational values. The result of F-test indicates that the distribution of selected stations is appropriate and sufficient to model the monthly E_{pan} and to investigate the climatic effects on the monthly E_{pan} modeling at the basin level. The Pearson correlation coefficient between data (precipitation, temperature, humidity at two timestamps, wind speed, and pan evaporation) was calculated at a significant level of 99% using SPSS software (IBM Corp. Released 2017), and the results showed a significant correlation between meteorological data and monthly E_{pan}. The statistical specifications of the data for the selected stations are given in Table 1.

Approaches

Wavelet

Wave-like functions that start at zero and end at zero oscillations form a data-driven method called "wavelet" (Danandeh Mehr et al., 2013; Ozger, 2010). Wavelet transformation is a mathematical tool used to extract information from different types of data (Partal & Kisi, 2007). Wavelet transformation does this by analyzing local deviations in time series. In this way, the method analyzes the data by extracting appropriate time fluctuations from unstable and temporary

Table 1 Statistical specifications of the study stations for 2005–2018 study years

Basin	Station	Longitude (°E)	Latitude (°N)	Variable Unit	P mm	T_{mean} °C	T_{min} °C	T_{max} °C	U km/hr	H_{18} %	H_{12} %	E_{pan} mm
Urmia Lake	Shahid Kazemi	46′ 34″	36′ 21″	Avg	29.52	13.25	6.33	20.18	3936.32	49.22	51.99	173.53
				S	31.82	9.36	8.01	10.77	2145.86	19.73	19.79	149.54
				C_V	1.08	0.71	1.27	0.53	0.55	0.40	0.38	0.86
	Pol-Sorkh	45′ 44″	36′ 46″	Avg	28.08	13.44	7.02	19.84	2443.80	67.02	62.28	143.09
				S	31.80	9.02	7.76	10.31	1243.79	11.02	13.90	120.43
				C_V	1.13	0.67	1.11	0.52	0.51	0.17	0.22	0.84
	Dashkhaneh	45′ 41″	37′ 01″	Avg	20.81	11.68	2.97	19.41	3056.16	64.38	64.84	119.51
				S	29.45	8.72	7.65	10.80	7345.75	11.36	8.90	100.64
				C_V	1.42	0.75	2.58	0.56	2.40	0.18	0.14	0.84
	Lighvan	46′ 26″	37′ 50″	Avg	32.72	7.76	1.44	13.91	2649.33	49.07	45.64	97.90
				S	21.62	8.26	7.25	9.44	752.51	13.39	8.58	84.52
				C_V	0.66	1.06	5.04	0.68	0.28	0.27	0.19	0.86
Gavkhouni	Ghaleh-Shahrokh	50′ 27″	32′ 40″	Avg	30.87	9.80	−0.61	19.87	3815.41	48.55	38.25	123.87
				S	37.60	8.78	7.27	10.82	1131.09	20.37	16.91	106.00
				C_V	1.22	0.90	−11.98	0.54	0.30	0.42	0.44	0.86
	Mahyar	51′ 48″	32′ 16″	Avg	11.25	16.81	9.15	24.45	2636.01	44.49	37.99	212.48
				S	15.75	8.87	8.58	9.21	643.91	15.75	13.44	143.60
				C_V	1.40	0.53	0.94	0.38	0.24	0.35	0.35	0.68
	Izad-Khast	52′ 08″	31′ 32″	Avg	11.43	13.90	6.59	20.93	6075.41	35.64	27.37	250.16
				S	17.00	8.19	7.56	9.08	1495.56	15.62	12.02	150.96
				C_V	1.49	0.59	1.15	0.43	0.25	0.44	0.44	0.60
	Morcheh-Khort	51′ 28″	33′ 06″	Avg	8.46	17.23	9.72	24.49	3108.11	35.63	27.86	264.64
				S	11.40	9.77	9.75	10.30	981.54	19.05	15.47	192.87
				C_V	1.35	0.57	1.00	0.42	0.32	0.54	0.56	0.73
	Meymeh	51′ 11″	33′ 26″	Avg	49.27	13.25	4.85	21.30	2759.17	34.64	26.17	180.77
				S	59.02	8.99	8.58	9.76	792.20	17.27	12.62	132.15
				C_V	1.20	0.68	1.77	0.46	0.29	0.50	0.48	0.73
	Pol-Zamankhan	50′ 54″	32′ 29″	Avg	28.58	13.30	3.38	22.92	1408.48	43.70	33.51	135.91
				S	37.75	8.00	6.58	9.75	540.96	13.04	11.41	97.54
				C_V	1.32	0.60	1.95	0.43	0.38	0.30	0.34	0.718

P, T_{mean}, T_{min}, T_{max}, H_{12}, H_{18}, U, and E_{pan} denote monthly precipitation, average temperature, minimum temperature, maximum temperatures, relative humidity at noon and 6 p.m., wind speed, and pan evaporation, respectively

Table 2 Results of wavelet analysis at level 1 for 2005–2018 study years

Basin	Station	Sub-series		SC -	P mm	T_{mean} °C	T_{min} °C	T_{max} °C	W km/hr	H_{18} %	H_{12} %	E_{pan} mm
Urmia Lake	Shahid Kazemi	A1[a]	min	1.49	−18.76	−5.50	−11.16	0.17	−513.80	8.16	13.37	−27.24
			max	11.99	93.57	28.50	19.68	37.28	16,762.59	85.88	89.89	459.55
		D1[b]	min	−4.23	−54.25	−2.87	−3.02	−2.93	−12,666.57	−18.25	−14.24	−68.48
			max	4.08	63.25	3.30	3.63	3.40	12,955.41	16.73	16.02	82.73
	Pol-Sorkh	A1	min	1.49	−18.06	−3.85	−9.40	1.09	−415.23	36.17	31.50	−28.42
			max	11.99	90.60	27.90	19.05	37.46	4880.21	82.67	84.81	365.65
		D1	min	−4.23	−68.15	−4.20	−3.66	−4.78	−1202.55	−11.50	−14.87	−58.37
			max	4.08	73.09	4.44	3.88	5.04	1012.54	10.93	14.06	43.90
	Dashkhaneh	A1	min	1.49	−23.54	−6.21	−13.85	−1.34	−13,120.40	38.78	40.62	−29.95
			max	11.16	95.62	25.77	14.24	37.51	51,415.17	84.91	81.22	283.14
		D1	min	−4.23	−87.22	−4.43	−4.11	−4.81	−42,721.96	−19.88	−10.07	−62.61
			max	4.08	93.38	5.65	4.78	5.39	45,210.33	11.12	10.37	68.89
	Lighvan	A1	min	1.49	−6.66	−7.78	−14.19	−5.38	1529.36	3.47	29.42	−19.86
			max	11.99	94.03	21.66	13.90	29.44	4427.32	79.49	64.91	254.70
		D1	min	−4.23	−40.80	−7.11	−4.53	−3.37	−1039.65	−14.25	−14.70	−220.47
			max	4.08	43.38	3.92	3.78	4.33	1075.55	16.50	11.34	54.70
Gavkhouni	Ghaleh-Shahrokh	A1	min	1.49	−10.73	−12.16	−21.16	−3.15	1778.45	11.17	11.65	−11.12
			max	11.16	185.28	22.27	9.99	36.97	6330.49	87.14	79.64	388.27
		D1	min	−4.23	−70.22	−5.70	−3.54	−8.64	−1633.84	−27.75	−13.84	−38.82
			max	4.08	64.72	6.21	3.96	8.19	1356.80	27.00	16.11	34.77
	Mahyar	A1	min	0.99	−8.62	1.03	−6.88	7.34	1343.83	10.19	7.91	−7.90
			max	11.16	59.61	30.49	22.57	38.32	4135.06	78.73	66.87	480.79
		D1	min	−4.23	−29.61	−2.17	−1.92	−2.74	−1248.13	−15.29	−13.50	−52.55
			max	4.08	29.05	3.56	3.23	3.90	1302.35	16.44	14.54	64.38
	Izad-Khast	A1	min	1.49	−8.24	−0.68	−7.43	5.54	3595.34	13.41	8.43	−13.56
			max	11.99	73.52	26.95	18.99	34.94	9535.53	69.89	57.72	489.68

Table 2 (continued)

Basin	Station	Sub-series		SC -	P mm	T_{mean} °C	T_{min} °C	T_{max} °C	W km/hr	H_{18} %	H_{12} %	E_{pan} mm
		D1	min	−4.23	−28.62	−4.10	−2.85	−7.02	−4010.88	−13.05	−14.02	−90.04
			max	4.08	30.10	5.07	2.60	7.25	4371.79	10.14	12.26	91.68
	Morcheh-Khort	A1	min	0.99	−6.34	0.18	−6.58	6.66	1383.87	8.10	6.36	−39.90
			max	11.16	39.09	33.42	26.87	40.27	5031.51	77.31	69.22	648.06
		D1	min	−4.23	−23.66	−3.49	−2.99	−7.97	−1477.06	−18.39	−14.15	−118.81
			max	4.08	27.41	3.91	3.25	7.13	1464.47	17.59	14.50	122.84
	Meymeh	A1	min	1.49	−18.65	−2.88	−11.58	3.78	1219.28	10.35	8.94	−8.34
			max	11.16	195.34	26.88	18.87	36.23	4771.35	84.98	57.11	416.60
		D1	min	−4.23	−120.75	−5.44	−3.38	−8.68	−1271.59	−25.83	−12.90	−44.63
			max	4.08	116.34	6.17	3.03	8.58	1284.58	24.42	12.42	35.97
	Pol-Zamankhan	A1	min	1.49	−14.60	−2.00	−10.17	5.09	232.80	18.52	14.71	−24.59
			max	11.16	152.70	26.06	14.04	38.06	2801.62	69.60	59.89	292.89
		D1	min	−4.23	−89.06	−5.62	−3.51	−7.21	−794.97	−8.25	−9.39	−54.82
			max	4.08	103.43	6.50	3.90	7.59	1281.75	10.05	8.79	50.10

[a]Main decomposition

[b]Sub-decomposition or data noise

signals by analyzing the frequency content of the signals in time series and shapes (Montaseri et al., 2018). Thus, the wavelet can well decompose the low-frequency dynamics of unstable time series and is also strong in displaying noise, seasonal patterns, and variance change (Popoola, 2007). The functions in the wavelet can separate the data into different frequency values. Each component of the data evaluates by a resolution appropriate to its scale. The lower scales show details of the signal that change rapidly, while the higher scales show the slow frequency changes of the components (Kucuk & Agiralioglu, 2006).

In general, there are two types of continuous and discrete waveform transformation. Because hydrological data have discrete time steps, discrete-time signals are used. Disconnected wavelet transformation that breaks down original data into bands is effective in improving predictions. The number of decomposition levels named “L” can be determined with Eq. (1) (Montaseri et al., 2018; Nourani et al., 2009):

$$L = \mathrm{int}\left[\log(N)\right] \tag{1}$$

where N is the number of data (in current study: $N=168$). So, based on Eq. (1), the level of wavelet decomposition is one. Decomposition and approximation in wavelet, using one-dimensional Daubechies-4 (db4) at level 1 of each variable, is presented in two parts, namely A1 and D1. An approximation of the signal is represented with A1 (low-frequency section), and details of the signal are represented with D1 (high-frequency section).

ANNs

The ANNs, which are made to mimic the neural networks of the human brain, contain layers of interconnected neurons that receive a set of inputs and parameters including weights and thresholds (Ahmadaali et al. 2013; Huo et al., 2012). By applying mathematical functions, the outputs are generated as a set of activators (stimulus functions). The layered structure of ANNs includes an input, an output, and one or more hidden (intermediate) layers. Each layer is associated with the surrounding layers and contains several nodes, neurons, or units that are networked together at different weights. Input units are the initial part of the network that receives the information from outside of the network and helps the network to learn the patterns of this information and finally processes those (Perera et al., 2020a, b).

ANFIS

Fuzzy systems are capable of modeling vague and uncertain problems, such as hydrological events. The ANFIS model was introduced in 1993 that enabled the fuzzy system and neural network capability (Jang, 1993). ANFIS is a multilayer feed-forward network with an input–output structure that applies neural network learning algorithms and fuzzy logic to transport inputs to output (Zareie & Fakherifard, 2014). In a set of input/output data, membership function parameters of the toolbox function ANFIS have two states to learn from the given data for modeling in a fuzzy system, a back-propagation (BP) algorithm alone, and a BP algorithm in combination with a least-squares type of method. The approaches of Mamdani (Mamdani & Assilian, 1975) and Sugeno (Takagi & Sugeno, 1985) are two approaches for fuzzy inference systems. The difference between the two approaches is the output membership functions of each model that is fuzzy membership functions in the Mamdani approach and linear or constant functions in Sugeno’s approach. Using Sugeno’s approach is more compact and computationally efficient than Mamdani’s approach (Kisi et al., 2013). In Takagi–Sugeno FIS, one of the most accurate fuzzy models frequently is used (Takagi & Sugeno, 1985). An ANFIS structure consists of a set of Takagi–Sugeno-type fuzzy IF–THEN rules, which can be used to model and map input–output data. In a fuzzy inference system, Sugeno’s first-order fuzzy model, which contains two IF–THEN fuzzy laws, is as follows:

$$\text{If } x \text{ is } A_1 \text{ and } y \text{ is } B_1, \text{ then } f_1 = p_1x + q_1y + r_1 \tag{2}$$

$$\text{If } x \text{ is } A_2 \text{and } y \text{ is } B_2, \text{ then } f_2 = p_2x + q_2y + r_2 \tag{3}$$

where x and y are the inputs and f is the output of the model, A_1, A_2, and B_1, B_2 denote the membership values of x and y as input variables, respectively; p_1, q_1, r_1, and p_2, q_2, r_2 are the parameters of the output functions f_1 and f_2, respectively.

GEP

In GEP, similar to the genetic algorithm (GA), linear and simple chromosomes of constant length are combined. Also, like decomposition trees in genetic programming (GP), there are branch structures with different sizes and shapes. GEP, by using a genetic algorithm, selects a population of individuals according to their suitability and applies genetic changes using one or more genetic operators. Thus, GEP is an extension of GP, which is also a branch of the GA (Holland, 1975).

The first step in the GEP algorithm is to generate a primary population of solutions. This can be done through a random process or by using some information about the problem. The chromosomes are then shown in tree expressions (ETs), which are also evaluated according to a fit function to determine the appropriateness of a solution in the problem area. The fitting function is usually evaluated by processing several sample problem targets, also called fitting items. If a satisfactory quality is found in a solution or a certain number of generations are achieved, evolution will stop and the best solution will be found. On the other hand, if the conditions are not met, the best solution will be kept from the present generation (meaning elite selection) and the rest of the solutions will be left to the selection process. Selecting the role of survival is worthwhile, and based on that, the best parents are chosen among the best people to produce the next generation. These processes are repeated over and over until it is expected that the quality of the generated population is improved. Figure 2 shows the structure of single approaches (ANN, ANFIS, and GEP).

Due to the high similarity of DWT wavelet tool (db4) trend to the current study dataset, a db4 approach was selected. The db4 mother wavelet trend is shown in Fig. 3. The ability of the wavelet method has led to the development of wavelet hybrids with other methods. In this study, the wavelet was hybridized with all three methods, ANN, GEP, and ANFIS and their results were compared. Figure 3 shows the used wavelet-hybrid approaches (WANN, WANFIS, and WGEP) structure, and Fig. 4 shows the stages of current research.

Calculating model performance

The performance accuracy of the models in the present study was evaluated with four criteria, namely squared correlation coefficient (R^2), root mean square error (RMSE), mean absolute error (MAE), and Nash–Sutcliffe efficiency (NSE) as follows:

$$R^2 = \left[\frac{\sum_{i=1}^{N} \left(E_{panio} - \overline{E_{pan_o}}\right)\left(E_{panie} - \overline{E_{pan_e}}\right)}{\sqrt{\sum_{i=1}^{N} \left(E_{panio} - \overline{E_{pan_o}}\right)^2 \sum_{i=1}^{N} \left(E_{panie} - \overline{E_{pan_e}}\right)^2}} \right]^2 \tag{4}$$

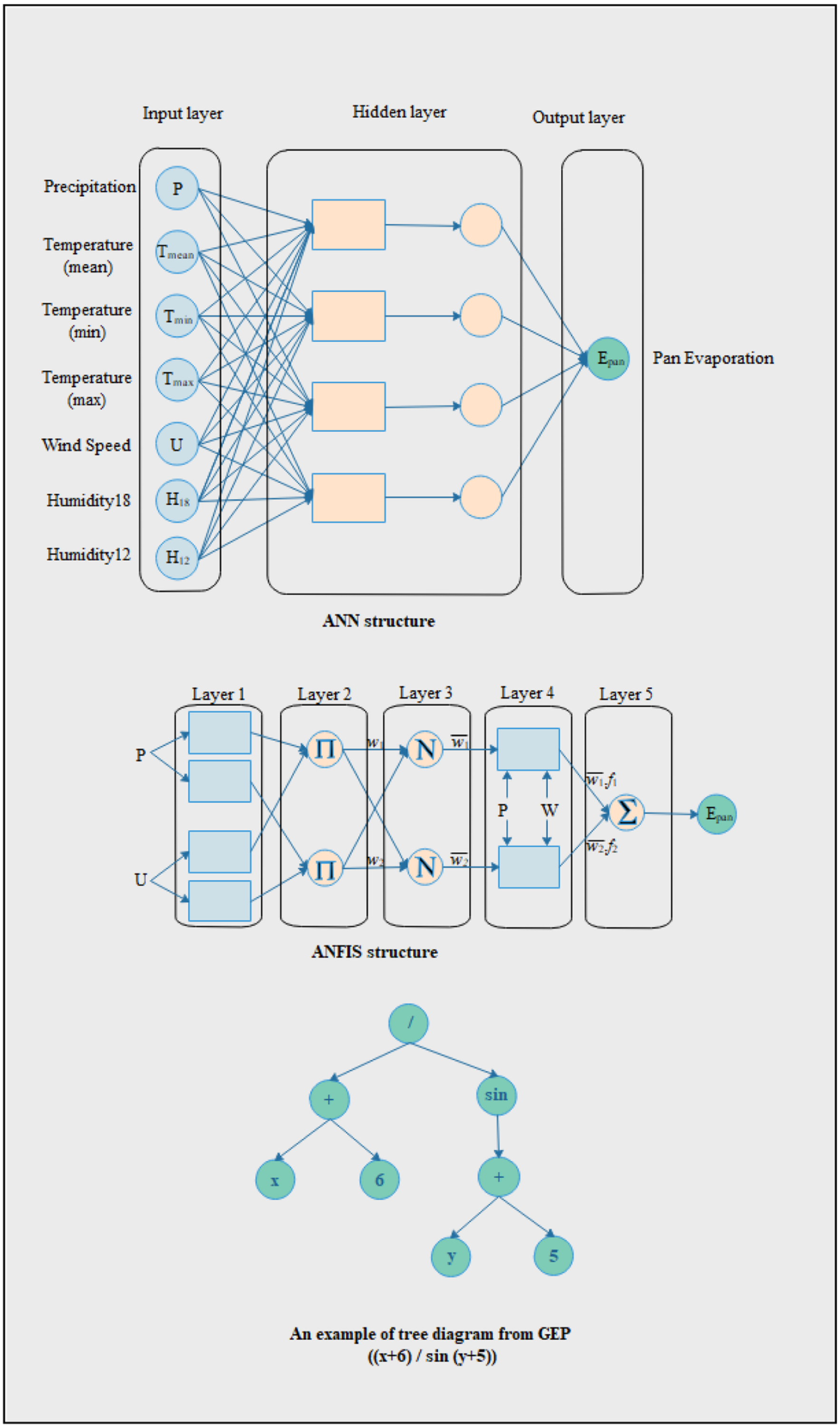
$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(E_{panio} - E_{panie}\right)^2} \tag{5}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} \left|E_{panio} - E_{panie}\right|^2 \tag{6}$$

$$NSE = 1 - \frac{\left(E_{panio} - E_{panie}\right)^2}{\left(E_{panio} - \bar{E}_{panie}\right)^2} \tag{7}$$

where E_{panio} and E_{panie} are the observed and estimated E_{pan} values for ith time step; N is the total number of observations; and $\overline{E_{pan_o}}$ and $\overline{E_{pan_e}}$ are the mean of observed and estimated E_{pan}. A positive R^2 shows that the correlation between observed and estimated values has a positive slope. The RMSE value represents the standard deviation of the estimating error of residuals; its range of variation is from zero to large positive values which zero for perfect estimates and large positive values for poor estimates (Ashrafzadeh et al., 2019; Ghavidel & Montaseri, 2014; Hyndman & Koehler, 2006; ZamanZad-Ghavidel et al., 2020). Also, MAE varies from zero for perfect estimates to large positive values for poor estimates. NSE efficiency can range from large negative values to 1. An efficiency of 1 (NSE = 1) corresponds to a perfect match of estimated and observed data. To train and test the models, 70% of the data (2005–2014) and 30% (2015–2018) were used, respectively. The training and testing ratio of dataset was selected based on trial and error. In current study, the data were normalized using Eq. 8.

Fig. 2 Used single approach (ANN, ANFIS, and GEP) structure



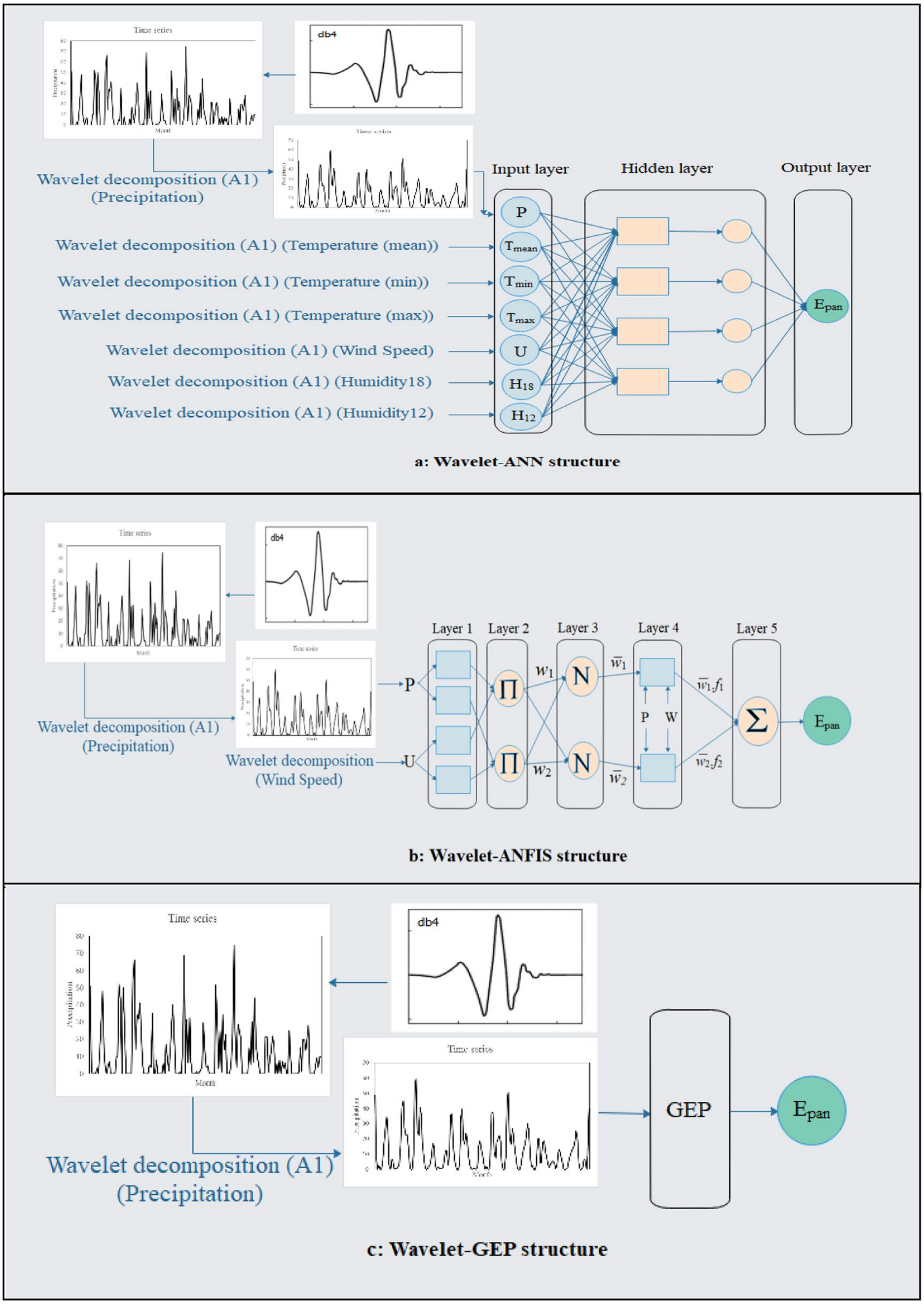


Fig. 3 Used wavelet-hybrid approach (**a** WANN, **b** WANFIS, and **c** WGEP) structure

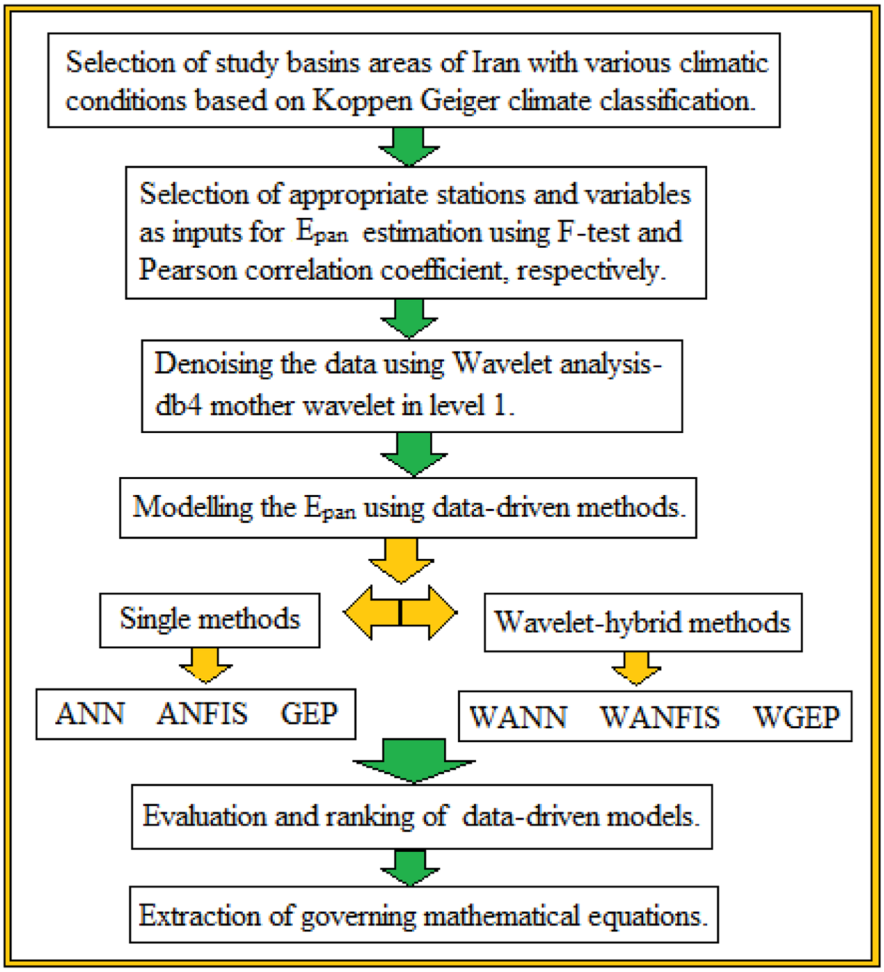


Fig. 4 Stages of current research

$$X_N = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{8}$$

where X_N, X, X_{min}, and X_{max} denote the normalize data, observed data, minimum and maximum amount of observed data, respectively.

Results and discussion

Wavelet theory and analysis

Decomposing the data into its sub-series was done by using db4 wavelet analysis. Prospering utilization of the db4 for data information in the hydrology and meteorology field has been offered by many researchers recently (Araghi et al., 2020; Montaseri et al., 2018; Seo et al., 2015). The db4 wavelet decomposed at level 1 for P, T_{mean}, T_{min}, T_{max}, U, H_{18}, H_{12}, E_{pan}, and seasonal coefficient variables (SC, with the values of 1 to 12 for September to October, respectively). Table 2 illustrates two series of obtained results based on wavelet analysis. The A1 series is known as the main decomposition and the D1 series as the sub-decomposition or data noise. Removal of the existing hidden noises in meteorological and hydrological data that affect the model results can be a powerful tool to increase modeling performance. Both measurement errors and uncertainty in hydrological phenomena are detected in the noise errors.

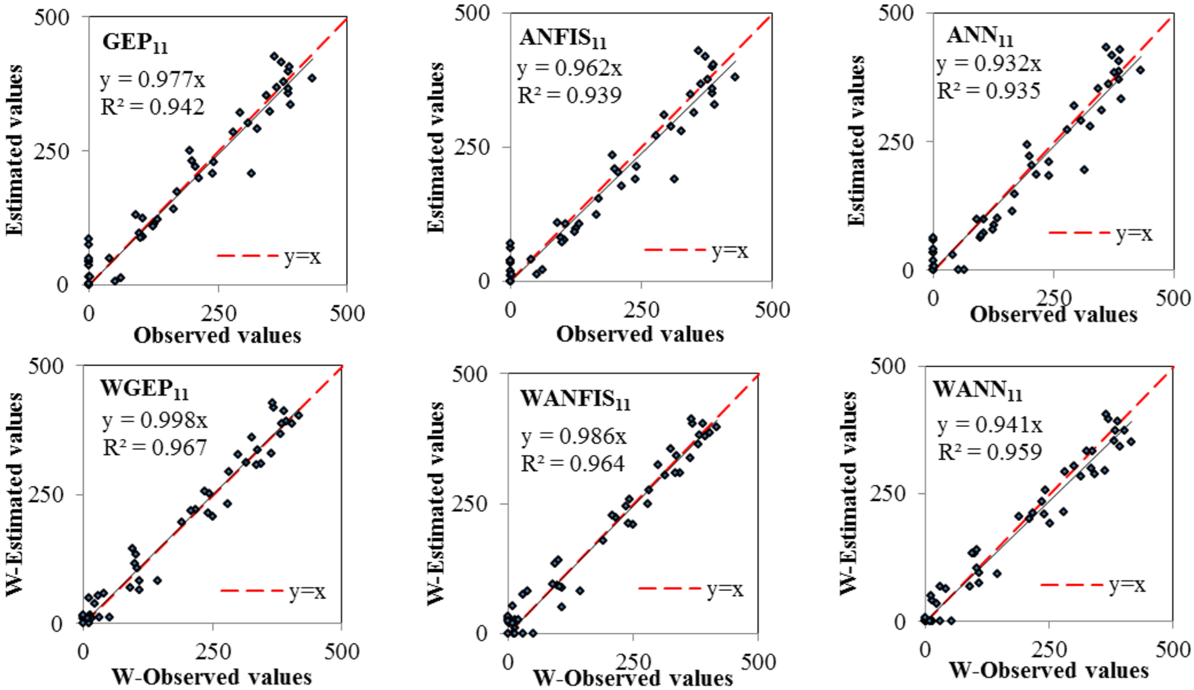


Fig. 5 The observed and estimated monthly E_{pan} (mm) values with the ANN, ANFIS, and GEP and their wavelet-hybrid (W-) models for Shahid Kazemi station and testing section

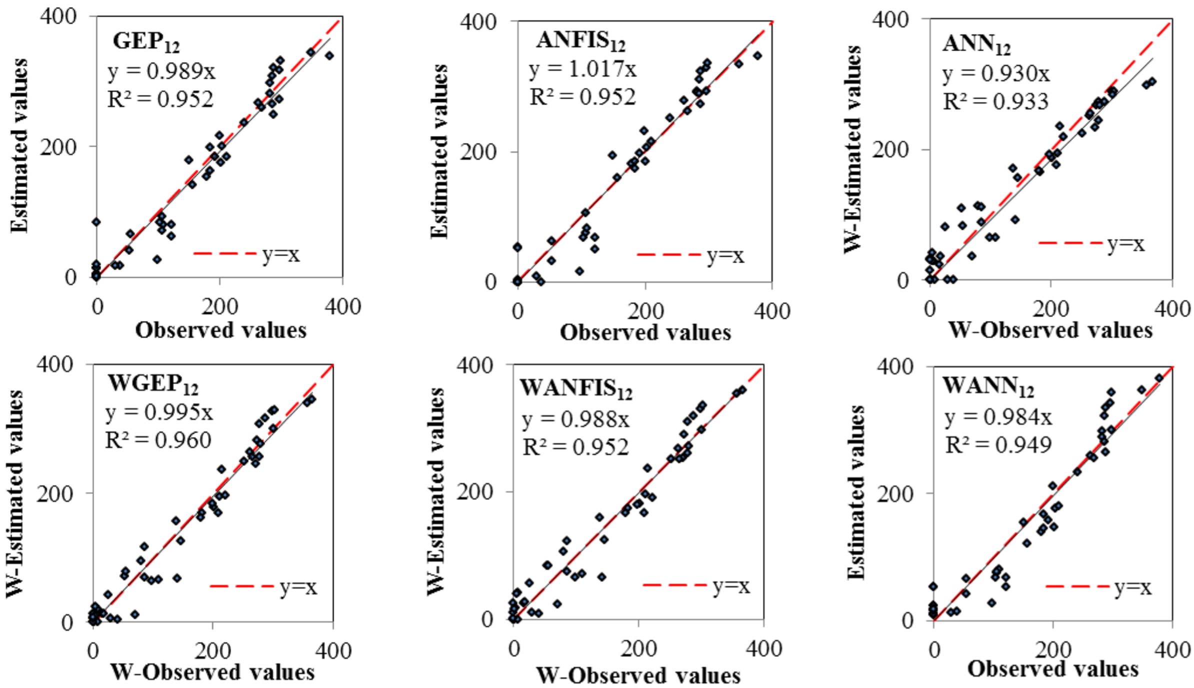


Fig. 6 The observed and estimated monthly E_{pan} (mm) values with the ANN, ANFIS, and GEP and their wavelet-hybrid (W-) models for Pol-Sorkh station and testing section

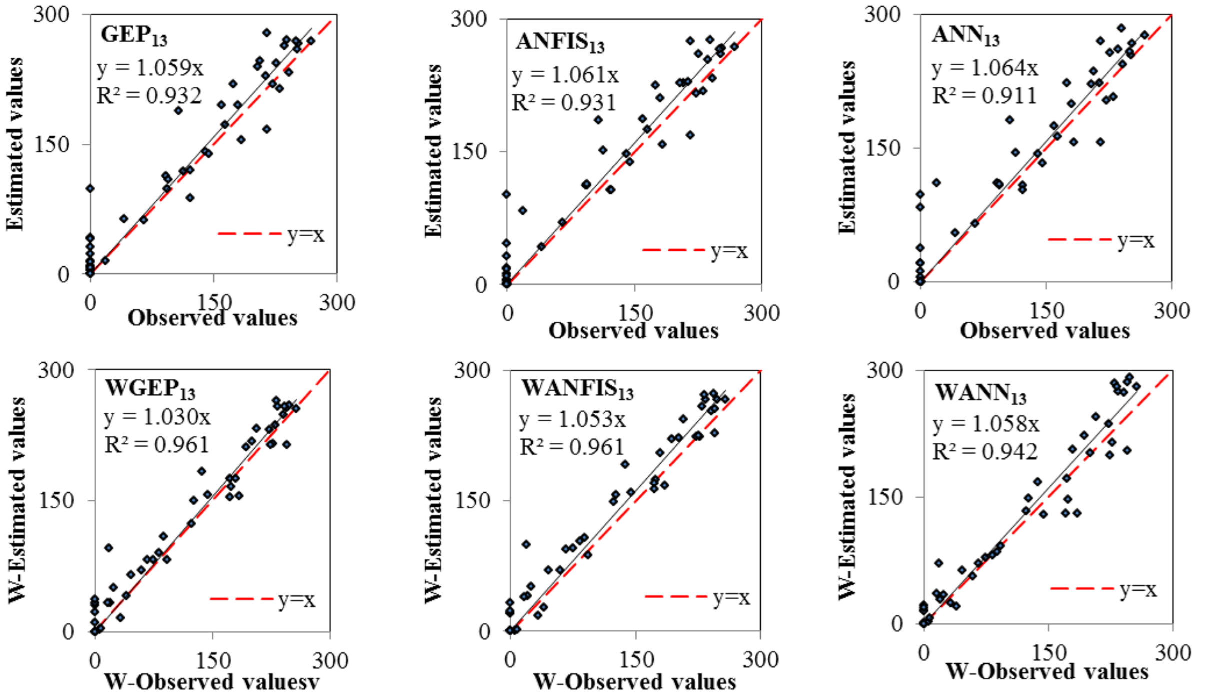


Fig. 7 The observed and estimated monthly E_{pan} (mm) values with the ANN, ANFIS, and GEP and their wavelet-hybrid (W-) models for Dashkhaneh station and testing section

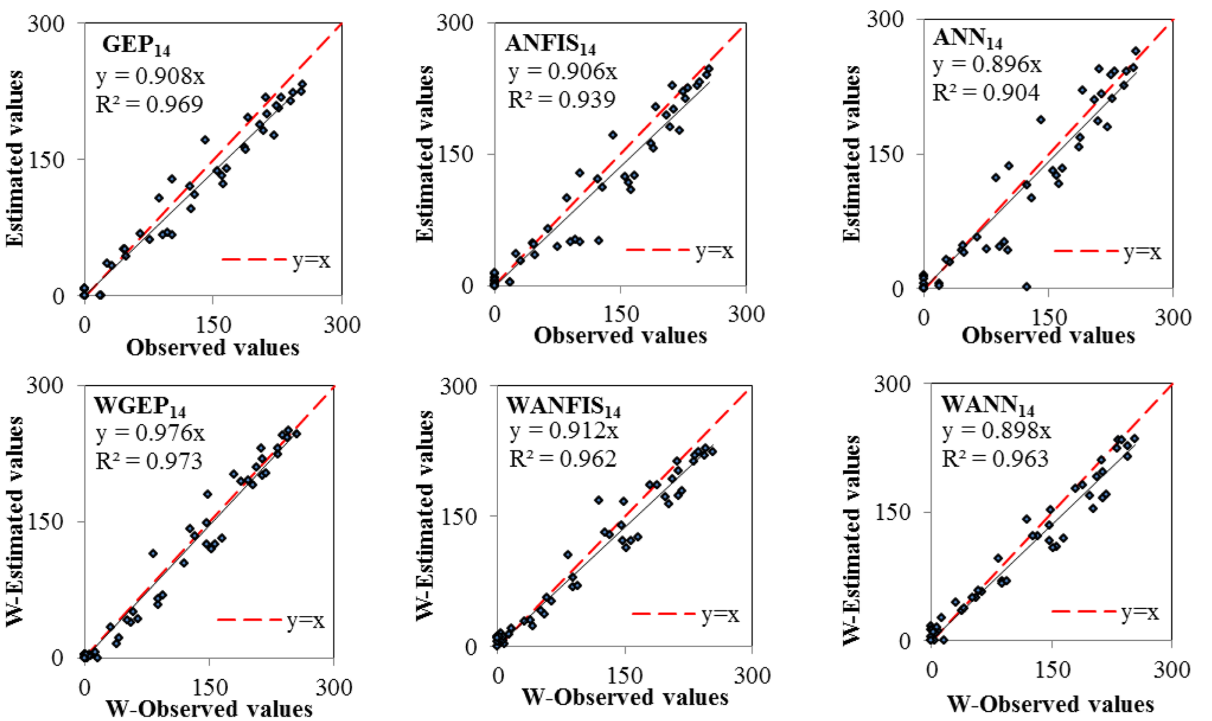


Fig. 8 The observed and estimated monthly E_{pan} (mm) values with the ANN, ANFIS, and GEP and their wavelet-hybrid (W-) models for Lighvan station and testing section

Single models values for estimating E_{pan} variable

To estimate monthly E_{pan}, three single data-driven approaches, namely ANN_{ij}, $ANFIS_{ij}$, and GEP_{ij} ($i = 1$ and 2 denote the number of basins for Urmia Lake and Gavkhouni, respectively, $j = 1,.., 4$ and $1,.., 6$ denote the number of selected stations in Urmia Lake and Gavkhouni basins, respectively), were used. The ANN-Levenberg–Marquardt (LM) backpropagation algorithm with one hidden layer was used. There is no special rule for determining the hidden nodes' number of ANN and radii values of the ANFIS models. The final values of them were determined through the trial and error process (were selected between 1 and 10 neurons number for ANN models). The results of optimal single ANN, ANFIS, and GEP models by R^2, RMSE, MAE, and NSE values and specifications of ANN (the number of neurons and activation functions of hidden and output layers) and ANFIS (radii values) models are listed in Tables 3, 4 for Urmia Lake and Gavkhouni basins, respectively. The optimal number of hidden layer neurons of models (were selected between 1 and 10 neurons number for ANN models) for ANN_{11} to ANN_{14} and ANN_{21} to ANN_{26} were varied from 2 to 5. The activation functions of output nodes were determined purelin or tansig or logsig for the ANN estimated models of stations in Urmia Lake and Gavkhouni basins with various climatic conditions. For example, the activation functions of hidden nodes of ANN_{11} to ANN_{14} models were tansig-purelin, logsig-purelin, tansig-purelin, and tansig-purelin for Urmia Lake basin. The radii values of $ANFIS_{11}$ to $ANFIS_{14}$ and $ANFIS_{21}$ to $ANFIS_{26}$ models were varied from 0.22 to 0.32 and 0.18 to 0.36, respectively.

Table 3 The results of optimal ANN, ANFIS, and GEP models and their wavelet-hybrid for Urmia Lake basin ($i = 1$) for 2005–2018 study years and testing section

J	Stations	Approach	Models	R^2	RMSE (mm)	MAE (mm)	NSE
1	Shahid Kazemi	Single	GEP_{11}	0.942	35.790	26.753	0.941
			$ANFIS_{11}$ (0.32)[a]	0.939	37.110	28.620	0.937
			ANN_{11} (LM, tansig-pureline-3)[b]	0.935	38.452	30.033	0.932
		Hybrid	$WGEP_{11}$	0.967	26.727	20.841	0.967
			$WANFIS_{11}$ (0.38)	0.964	27.690	23.017	0.964
			$WANN_{11}$ (LM, tansig-tansig-3)	0.959	31.433	24.537	0.954
2	Pol-Sorkh	Single	GEP_{12}	0.952	26.360	19.649	0.950
			$ANFIS_{12}$ (0.28)	0.952	27.196	21.203	0.947
			ANN_{12} (LM, logsig-pureline-2)	0.933	31.345	25.671	0.929
		Hybrid	$WGEP_{12}$	0.960	23.928	18.956	0.958
			$WANFIS_{12}$ (0.31)	0.952	25.580	20.974	0.952
			$WANN_{12}$ (LM, tansig-pureline-5)	0.949	28.420	23.158	0.940
3	Dashkhaneh	Single	GEP_{13}	0.932	29.291	20.403	0.910
			$ANFIS_{13}$ (0.22)	0.931	29.984	20.791	0.906
			ANN_{13} (LM, tansig-pureline-3)	0.911	32.744	21.677	0.888
		Hybrid	$WGEP_{13}$	0.960	21.055	15.638	0.951
			$WANFIS_{13}$ (0.35)	0.961	23.681	18.128	0.937
			$WANN_{13}$ (LM, tansig-tansig-4)	0.942	26.687	20.020	0.921
4	Lighvan	Single	GEP_{14}	0.969	19.467	15.432	0.952
			$ANFIS_{14}$ (0.32)	0.939	24.791	18.016	0.922
			ANN_{14} (LM, tansig-pureline-2)	0.904	29.397	19.877	0.891
		Hybrid	$WGEP_{14}$	0.973	15.839	11.923	0.968
			$WANFIS_{14}$(0.41)	0.962	20.035	15.396	0.948
			$WANN_{14}$ (LM, tansig-pureline-2)	0.963	21.092	15.621	0.942

[a]Radii values

[b]Training algorithm, hidden layer activation functions, output layer activation functions, number of neurons in hidden layer

Table 4 The results of optimal ANN, ANFIS, and GEP models and their wavelet-hybrid for Gavkhouni basin ($i=2$) for 2005–2018 study years and testing section

j	Stations	Approach	Models	R^2	RMSE (mm)	MAE (mm)	NSE
1	Ghaleh-Shahrokh	Single	GEP_{21}	0.942	26.302	21.643	0.951
			$ANFIS_{21}$(0.18)	0.924	30.492	24.448	0.921
			ANN_{21}(LM, tansig-tansig-2)	0.920	32.495	26.561	0.911
		Hybrid	$WGEP_{21}$	0.954	23.782	19.074	0.984
			$WANFIS_{21}$(0.24)	0.947	25.972	20.884	0.941
			$WANN_{21}$(LM, tansig-pureline-3)	0.939	27.164	20.249	0.936
2	Mahyar	Single	GEP_{22}	0.940	37.806	31.951	0.922
			$ANFIS_{22}$(0.28)	0.937	51.507	41.147	0.856
			ANN_{22}(LM, tansig-tansig-3)	0.924	54.145	42.112	0.840
		Hybrid	$WGEP_{22}$	0.978	28.915	24.488	0.954
			$WANFIS_{22}$(0.32)	0.972	33.588	28.965	0.937
			$WANN_{22}$(LM, tansig-pureline-2)	0.949	46.173	38.292	0.881
3	Izad-Khast	Single	GEP_{23}	0.963	29.845	25.942	0.946
			$ANFIS_{23}$(0.32)	0.954	34.518	27.962	0.928
			ANN_{23}(LM, tansig-pureline-5)	0.939	36.745	28.801	0.918
		Hybrid	$WGEP_{23}$	0.972	26.319	22.100	0.957
			$WANFIS_{23}$(0.35)	0.969	29.882	24.706	0.945
			$WANN_{23}$(LM, tansig-tansig-3)	0.952	33.197	26.551	0.932
4	Morcheh-Khort	Single	GEP_{24}	0.895	80.774	61.570	0.792
			$ANFIS_{24}$(0.36)	0.874	83.146	62.396	0.758
			ANN_{24}(LM, logsig-pureline-3)	0.858	85.237	63.048	0.709
		Hybrid	$WGEP_{24}$	0.921	70.318	57.815	0.796
			$WANFIS_{24}$(0.28)	0.917	74.220	58.455	0.769
			$WANN_{24}$(LM, tansig-tansig-4)	0.904	81.576	63.035	0.731
5	Meymeh	Single	GEP_{25}	0.971	23.493	19.728	0.968
			$ANFIS_{25}$(0.21)	0.968	25.281	21.565	0.963
			ANN_{25}(LM, tansig-pureline-5)	0.951	30.635	24.443	0.945
		Hybrid	$WGEP_{25}$	0.979	20.651	17.295	0.975
			$WANFIS_{25}$(0.39)	0.970	23.997	18.402	0.966
			$WANN_{25}$(LM, tansig-logsig-2)	0.954	29.787	18.653	0.947
6	Pol-Zamankhan	Single	GEP_{26}	0.934	28.958	20.295	0.920
			$ANFIS_{26}$(0.33)	0.919	31.122	21.904	0.908
			ANN_{26}(LM, logsig-pureline-3)	0.909	33.169	25.663	0.895
		Hybrid	$WGEP_{26}$	0.959	25.881	19.784	0.934
			$WANFIS_{26}$(0.27)	0.952	27.897	21.514	0.909
			$WANN_{26}$(LM, tansig-tansig-5)	0.935	32.022	24.422	0.898

The values of R^2, RMSE, MAE, and NSE for stations of Urmia Lake basin with Dsa climate classification corresponding to the ANN_{11} to ANN_{14} models were varied from 0.904 to 0.935, 29.397 to 38.452, 19.877 to 30.033, and 0.891 to 0.932, respectively. The values of R^2, RMSE, MAE, and NSE for stations of Gavkhouni basin with Bsk climate classification corresponding to the $ANFIS_{21}$ to $ANFIS_{26}$ models were varied 0.874 to 0.968, 25.281 to 83.146, 21.565 to 62.396, and 0.758 to 0.963, respectively. Concerning the GEP models, the root relative squared error (RRSE) was selected as the fitness function with pressure tree. The values of R^2 and RMSE for GEP_{11} to GEP_{14} and GEP_{21} to GEP_{26} models of E_{pan}

estimations for the selected stations of Urmia Lake and Gavkhouni basins were varied from R^2: 0.932 to 0.969 and RMSE (mm): 19.467 to 35.790 and R^2: 0.895 to 0.971 and RMSE (mm): 23.493 to 80.774, respectively. Tables 3, 4 mention the results of R^2, RMSE, MAE, and NSE values for ANN, ANFIS, and GEP models to estimate monthly E_{pan} variable for selected stations of both Urmia Lake and Gavkhouni basins with Dsa and Bsk climate types. Result of analysis showed that the artificial intelligence models are highly dependent on the training data (Chia et al., 2020).

Wavelet-hybrid models values for estimating E_{pan} variable

For improving the impact of single data-driven models and development of the hybrid-wavelet ANN_{ij}, $ANFIS_{ij}$, and GEP_{ij} ($WANN_{ij}$, $WANFIS_{ij}$, and $WGEP_{ij}$) estimation models via wavelet tools, the first stage is to decompose the selected climatology variables, into the subseries of details and approximations (A1 and D1 series) using a DWT wavelet tool (db4) that were written in MATLAB by researchers. To build and develop the hybrid-wavelet E_{pan} estimation models (WANN, WANFIS, and WGEP), D1 decomposed subseries were recognized as noise and removed from the models. After noise reduction, A1 decomposed subseries with positive values for positively selected variables were estimated separately with ANN, ANFIS, and GEP models.

The results of optimal wavelet-hybrid WANN, WANFIS, and WGEP models by R^2, RMSE, MAE, and NSE values and specifications of models for selected stations of both Urmia Lake and Gavkhouni basins with Dsa and Bsk climate classification are listed in Tables 3, 4, respectively. The optimal number of hidden layer neurons of models (were selected between 1 and 10 neurons number for WANN models) for $WANN_{11}$ to $WANN_{14}$ and $WANN_{21}$ to $WANN_{26}$ were varied from 2 to 5. The radii values of $WANFIS_{11}$ to $WANFIS_{14}$ and $WANFIS_{21}$ to $WANFIS_{26}$ models were varied from 0.31 to 0.41 and 0.24 to 0.39, respectively.

The values of R^2 and RMSE for stations of Urmia Lake and Gavkhouni basins with Dsa and Bsk climate classification corresponding to the $WANN_{11}$ to $WANN_{14}$ and $WANN_{21}$ to $WANN_{26}$ models were varied from R^2: 0.942 to 0.963 and RMSE (mm): 21.092 to 31.433, R^2: 0.904 to 0.954 and RMSE (mm): 27.164 to 81.576, respectively. On the other hand, the values of R^2 and RMSE corresponding to the $WANFIS_{11}$ to $WANFIS_{14}$ models were varied from R^2: 0.952 to 0.964 and RMSE (mm): 20.035 to 27.690 for stations of Urmia Lake basin, respectively. The values of R^2 and RMSE for $WGEP_{11}$ to $WGEP_{14}$ and $WGEP_{21}$ to $WGEP_{26}$ models of E_{pan} estimates for the selected stations of Urmia Lake and Gavkhouni basins were varied from R^2: 0.960 to 0.973 and RMSE (mm): 15.839 to 26.727 and R^2: 0.921 to 0.979 and RMSE (mm): 20.651 to 70.318, respectively.

Figures 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 illustrate the main and estimated monthly E_{pan} values for ANN, ANFIS, and GEP models and their wavelet-hybrid models for study stations located in Dsa and Bsk climate classification, respectively. The R^2 values for single and wavelet-hybrid data-driven models are close to 1, with the quality relations being $R^2_{GEP} > R^2_{ANFIS} > R^2_{ANN}$ and $R^2_{WGEP} > R^2_{WANFIS} > R^2_{WANN}$ for all stations studied basins. For example, for Dashkhaneh station in the Urmia Lake basin, the percentage of WGEP, WANFIS, and WANN performance improvement compared to GEP, ANFIS, and ANN is 28%, 21%, and 18%, respectively. Also, the percentage of WGEP, WANFIS, and WANN performance improvement compared to GEP, ANFIS, and ANN is 24%, 35%, and 15%, respectively, for Mahyar station in the Gavkhouni basin. The proximity to the value of one of a coefficient of the fitted equation ($y = ax$) to the observed and estimated values indicates better performance of the model. As Figs. 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 display, $WGEP_{11}$ to $WGEP_{14}$ and $WGEP_{21}$ to $WGEP_{26}$ models for stations of Urmia Lake and Gavkhouni basins have the closest values to one, respectively.

The results of all applied models established that the ANFIS model exceeded the ANN models' performance and WANFIS model exceeded the WANNs; also, the GEP and WGEP models had a better performance for monthly E_{pan} estimation than two other single and wavelet-hybrid applied models for both basins, respectively.

The box-plots are a graphic for showing image information instead of estimated monthly E_{pan} data. One of the best ways to accurate distribution understanding of estimated data is the box diagram, which is based on the five values of "minimum",

Table 5 Developed GEP and WGEP mathematical equations of the E_{pan} estimation for 2005–2018 study years

Basin	Station		Equation
Urmia lake	Shahid Kazemi	GEP	$E_{\text{pan x,t}} = (((Tmean_{x,t}) \times (Tmean_{x,t} - T\min_{x,t})) + (Tmean_{x,t})) + ((\tan^{-1}(T\min_{x,t})) \times (Tmean_{x,t})) + (((Tmean_{x,t} - T\min_{x,t}) \times (\tan^{-1}(Tmean_{x,t}))) + Tmean_{x,t})$
		WGEP	$E_{\text{pan x,t}} = (Tmean_{x,t}) + \left(Tmean_{x,t} \times (8.38 \times (8.48 - T\max_{x,t}))^{6}\right) + (T\max_{x,t})$
	Pol-Sorkh	GEP	$E_{\text{pan x,t}} = \left((T\min_{x,t})^{2} + T\min_{x,t}\right) + ((SC_{x,t} + (2.69 \times Tmean_{x,t})) \times Sin(SC_{x,t})) + (Tmean_{x,t} + T\max_{x,t})$
		WGEP	$E_{\text{pan x,t}} = (T\min_{x,t} \times (\exp(\tan^{-1}((T\min_{x,t}) \times (\exp(H12_{x,t} + T\min_{x,t})))))) + (4.81 \times T\min_{x,t}) + \left(\left(\sqrt[2]{T\max_{x,t}}\right) \times Tmean_{x,t}\right)$
	Dashkhaneh	GEP	$E_{\text{pan x,t}} = (9.28 \times T\max_{x,t}) + ((Cos(Ln(SC_{x,t}))) \times (T\max_{x,t})) + ((\exp(\tan^{-1}(H18_{x,t} - 7.53))) - ((H18x_{x,t}) - ((SC_{x,t} + 8.53) - (Tmean_{x,t}))))$
		WGEP	$E_{\text{pan x,t}} = (7.26 \times T\max_{x,t}) + ((51.09 - H18_{x,t}) - SC_{x,t}) - T\min_{x,t} + 0.061 + 576Tmean_{x,t}/(1.74 + T\max_{x,t})$
	Lighvan	GEP	$E_{\text{pan x,t}} = (8.29 \times T\min_{x,t}) + \left(\frac{\tan^{-1}(-0.42 + (-0.70 \times Tmean_{x,t}))}{(-0.42SC_{x,t}/Tmean_{x,t})}\right) + (6.15 + T\max_{x,t})$
		WGEP	$E_{\text{pan x,t}} = (Tmean_{x,t} \times ((\exp(\tan^{-1}(Tmean_{x,t}))) \times ((\tan^{-1}(H18_{x,t})) + 0.65))) + (\sqrt[2]{P_{x,t}}) + (Tmean_{x,t} \times (\exp(\cos(\tan^{-1}((SC_{x,t} - 7.66) + (\sqrt[2]{H18_{x,t}}))))))$
Gavkhouni	Ghaleh-Shahrokh	GEP	$E_{\text{pan x,t}} = \left((T\max_{x,t} + T\min_{x,t}) \times \left(\sqrt[2]{(T\max_{x,t} + 7.73)}\right)\right) + ((Cos(Tmean_{x,t})) - T\min_{x,t}) + \left(\left(\left((T\max_{x,t})^{3} - (Ln \div (W_{x,t}))\right)(W_{x,t})\right) - (Cos(W_{x,t}))\right)$
		WGEP	$E_{\text{pan x,t}} = (\tan^{-1}(T\max_{x,t})) + ((((0.16 - T\min_{x,t}) - Tmean_{x,t}) - SC_{x,t}) + (Cos(2 \times P_{x,t}))) + \left(\left(\sqrt[2]{\left((T\max_{x,t}) \times \left(Ln(H12_{x,t})^{2}\right)\right)}\right) \times (Tmean_{x,t})\right)$
	Mahyar	GEP	$E_{\text{pan x,t}} = \left((T\min_{x,t})^{2} - (T\min_{x,t})\right) + (5069.59 + H12_{x,t}) + ((Sin(SC_{x,t})) \times (0.97 \times ((SC_{x,t} - T\max_{x,t}) - T\max_{x,t})))$
		WGEP	$E_{\text{pan x,t}} = \left(\left((Cos(-0.87 - T\max_{x,t})) - \left(\left(\sqrt[3]{W_{x,t}}\right) \times (0.087 - Tmean_{x,t})\right)\right) - (T\max_{x,t})\right) + \left(\left(\left(\left(\sqrt[2]{Tmean_{x,t}}\right) \times (-3.52 - SC_{x,t})\right) + (3.77)\right) + (T\max_{x,t})\right) + (Tmean_{x,t})$
	Izad-Khast	GEP	$E_{\text{pan x,t}} = ((Ln((1.65 \times P_{x,t}) + (T\max_{x,t}/SC_{x,t}))) + (7.35 \times (T\max_{x,t} + Tmean_{x,t}))) + \left(3.53 + \left(\left(\sqrt[3]{W_{x,t}}\right) - ((Tmean_{x,t} - 4.94) + (SC_{x,t} - T\min_{x,t}))\right)\right) + \left(2.57 \times \left(\left(\sqrt[3]{W_{x,t}}\right) - (H12_{x,t})\right)\right)$
		WGEP	$E_{\text{pan x,t}} = \left(\left(\left(\sqrt[2]{(7.52 \times Tmean_{x,t})}\right) \times (Cos(H12_{x,t}))\right) + 56.61\right) + ((13.91 \times Tmean_{x,t}) - ((-2.21 \times T\min_{x,t}) + (H12_{x,t} + SC_{x,t}))) + (T\max_{x,t})$
	Morcheh-Khort	GEP	$E_{\text{pan x,t}} = \left(\sqrt[3]{(Tmean_{x,t} \times (((T\min_{x,t} + 7.49) \times Tmean_{x,t}) \times (W_{x,t} - P_{x,t}))))}\right) + \left(Tmean_{x,t} + \left(-7.88 - \left(\sqrt[4]{(SC_{x,t} \times W_{x,t})}\right)\right)\right) + \left(\left(Sin\left(\sqrt[2]{SC_{x,t}}\right) \times H12_{x,t}\right) \times 2.98\right)$

Table 5 (continued)

Basin	Station		Equation
		WGEP	$E_{pan\ x,t} = (((2\times H12_{x,t}) + (T\min_{x,t} - 6.74)) + ((Cos(H12_{x,t})) - SC_{x,t}^{2})) + \left(\sqrt[3]{\left(T\max_{x,t}\times(((W_{x,t}+T\min_{x,t}) - 293.29)\times(Tmean_{x,t})^{2})\right)}\right) + (Ln(Sin(EXP(\tan^{-1}((Cos(T\max_{x,t}/T\min_{x,t})) - (H18_{x,t}\times Tmean_{x,t}))))))$
	Meymeh	GEP	$E_{pan\ x,t} = (6.76\times T\min_{x,t}) + (6.67\times T\max_{x,t}) + \left(T\min_{x,t}\times\left(Cos\left((Cos(P_{x,t}))\times\left(\sqrt[3]{T\min_{x,t}^{2}}\right)\right)\right)\right)$
		WGEP	$E_{pan\ x,t} = \left(\sqrt[3]{\left(((T\max_{x,t}+2.61)\times W_{x,t}) + ((T\min_{x,t} - P_{x,t})\times(SC_{x,t})^{2})\right)}\right) + \left((Sin(SC_{x,t}+T\min_{x,t})) + \left(\left(\sqrt[2]{T\max_{x,t}}\right)\times(2\times Tmean_{x,t})\right)\right) + (Tmean_{x,t} - SC_{x,t})$
	Pol-Zamankhan	GEP	$E_{pan\ x,t} = (H12_{x,t} + ((T\min_{x,t} - (SC_{x,t} - 4.52)) - (H18_{x,t} - Tmean_{x,t}))) + (H12_{x,t} + ((T\min_{x,t} - (SC_{x,t} - 4.52)) - (H18_{x,t} - Tmean_{x,t}))) + (5.62\times T\max_{x,t})$
		WGEP	$E_{pan\ x,t} = ((T\max_{x,t}\times Ln(T\max_{x,t}))\times(Sin(2.85/H18_{x,t}))) + (6.02\times T\min_{x,t}) + (5.04\times T\max_{x,t})$

"first quartile (0.25%)", "median (0.50%)", "third quartile (0.75%)", and "maximum" values. It is also showing the symmetry of the data on the efficiency of this graph. It is worth noting that the focus and even the sloping of the data are shown in this graph. Figure 15 displays the boxplots of estimated

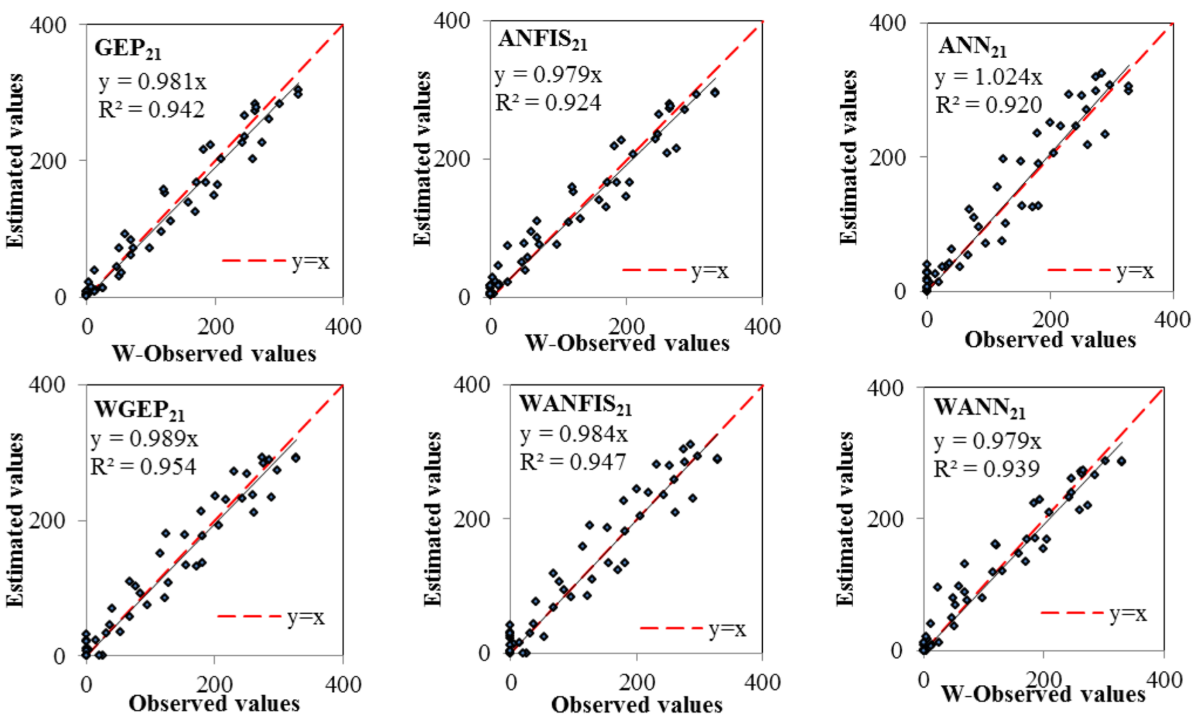


Fig. 9 The observed and estimated monthly E_{pan} (mm) values with the ANN, ANFIS, and GEP and their wavelet-hybrid (W-) models for Ghaleh-Shahrokh station and testing section

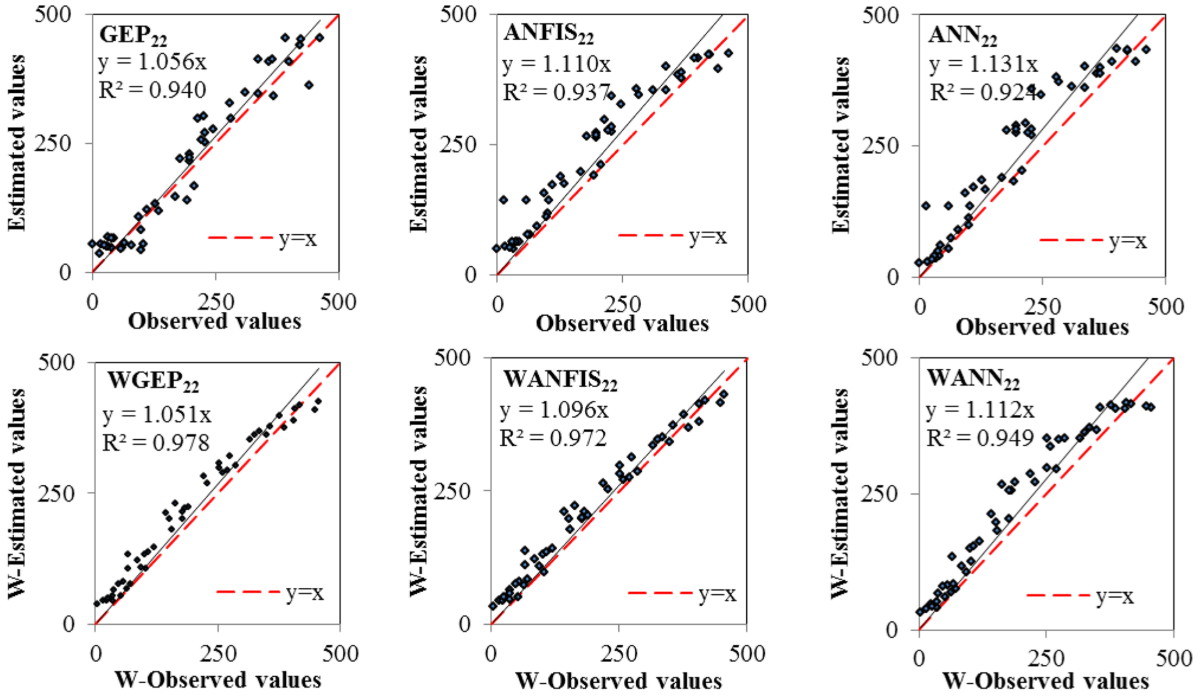


Fig. 10 The observed and estimated monthly E_{pan} (mm) values with the ANN, ANFIS, and GEP and their wavelet-hybrid (W-) models for Mahyar station and testing section

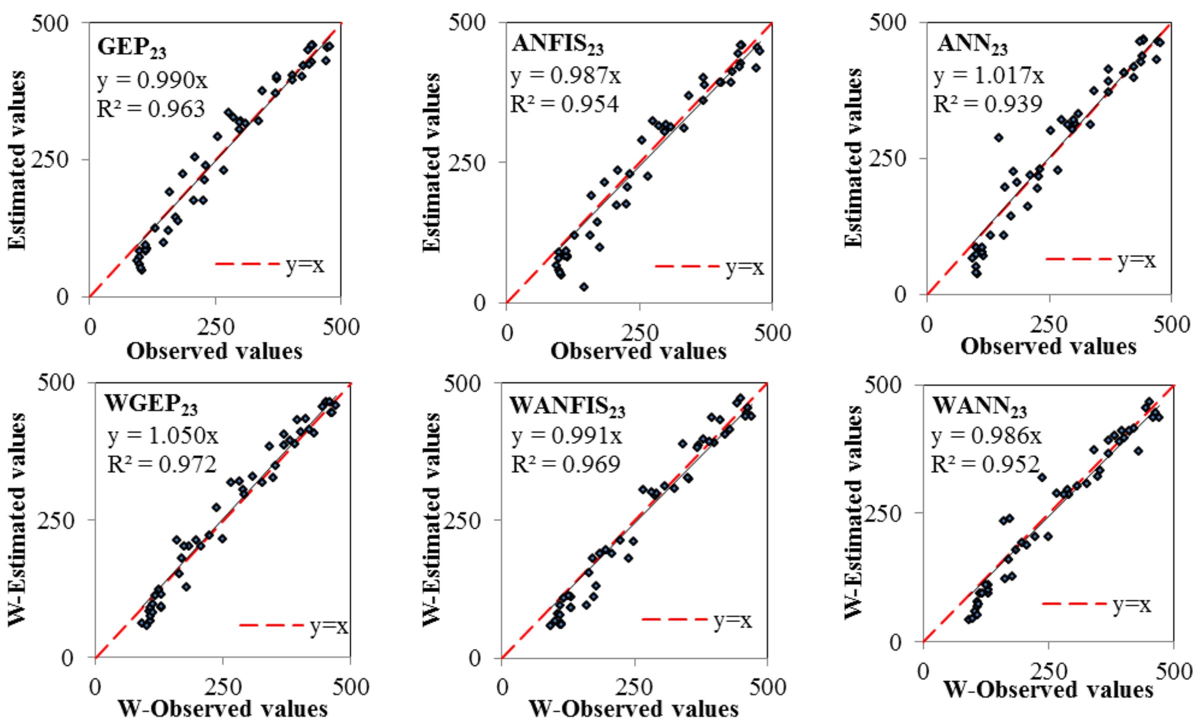


Fig. 11 The observed and estimated monthly E_{pan} (mm) values with the ANN, ANFIS, and GEP and their wavelet-hybrid (W-) models for Izad-Khast station and testing section

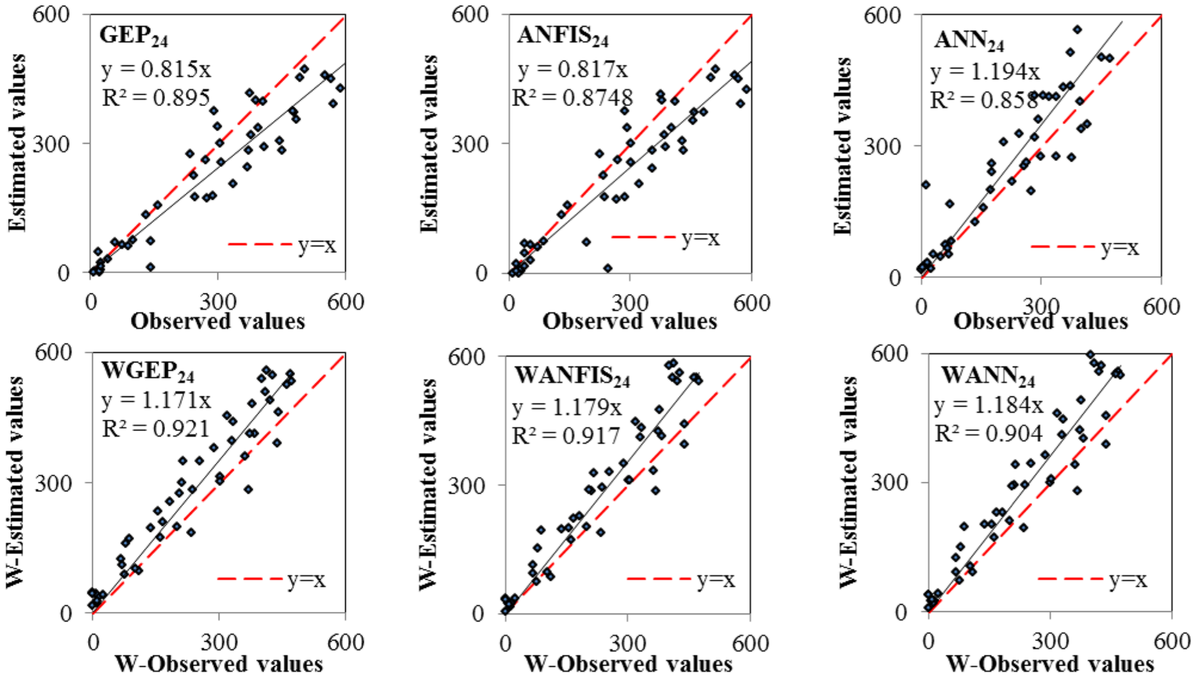


Fig. 12 The observed and estimated monthly E_{pan} (mm) values with the ANN, ANFIS, and GEP and their wavelet-hybrid (W-) models for Morcheh-Khort station and testing section

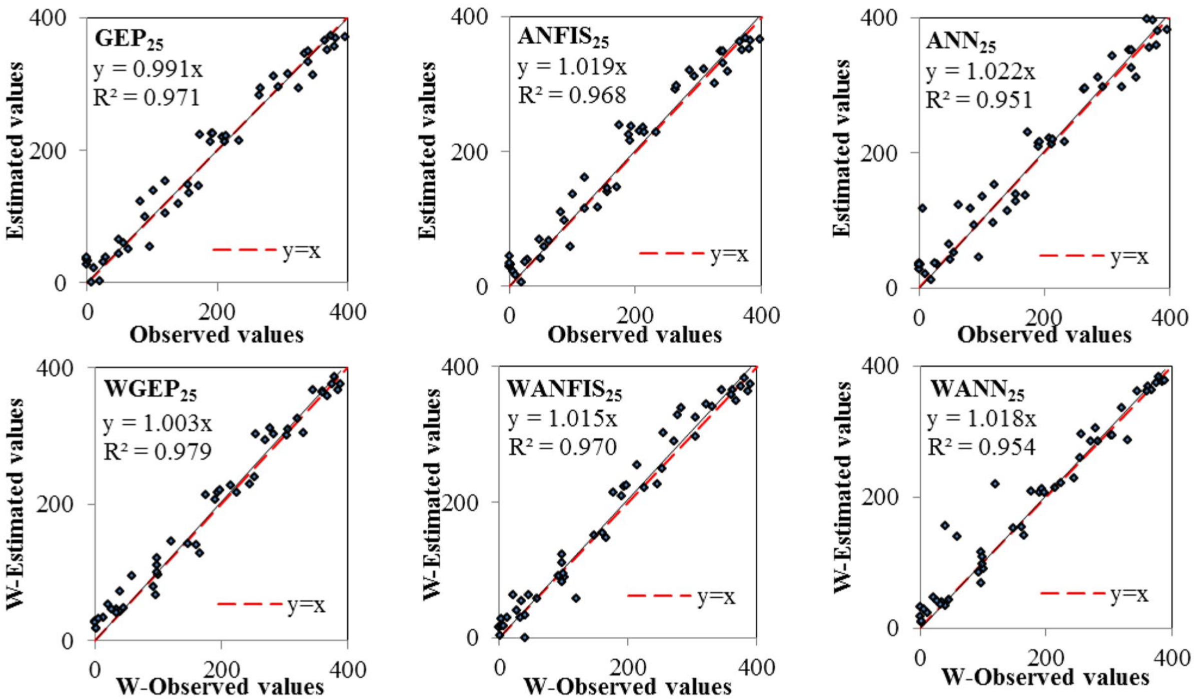


Fig. 13 The observed and estimated monthly E_{pan} (mm) values with the ANN, ANFIS, and GEP and their Wavelet-hybrid (W-) models for Meymeh station and testing section

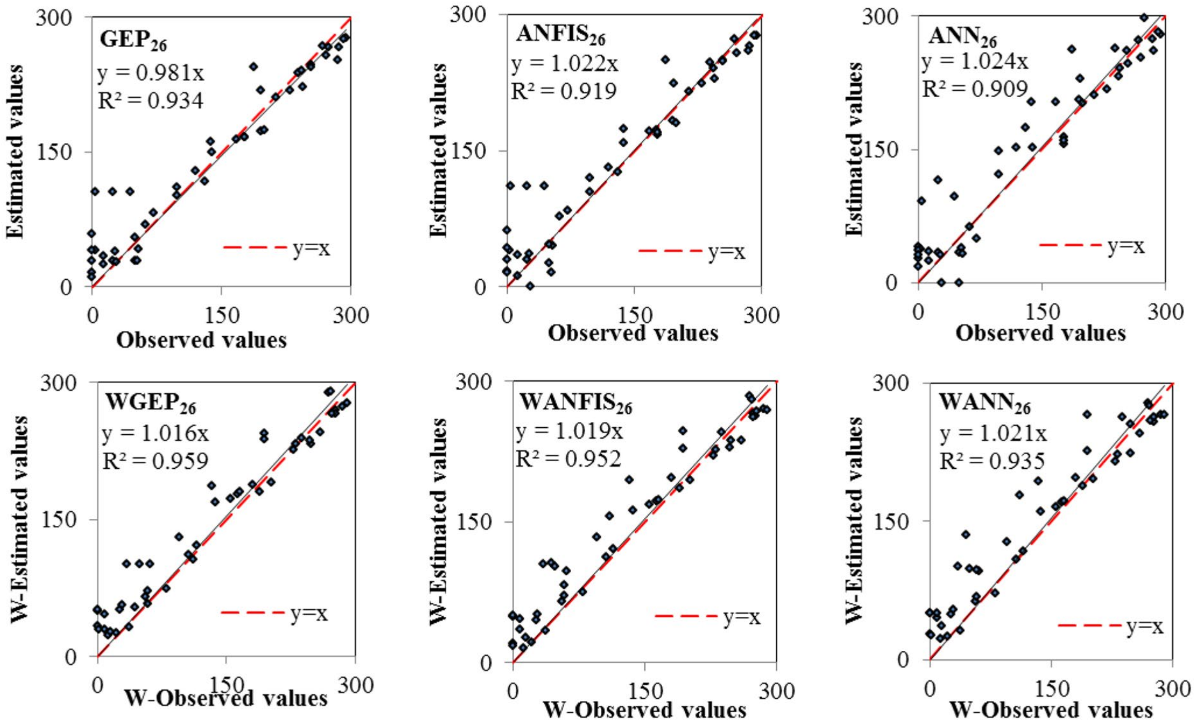


Fig. 14 The observed and estimated monthly E_{pan} (mm) values with the ANN, ANFIS, and GEP and their wavelet-hybrid (W-) models for Pol-Zamankhan station and testing section

E_{pan} values for Urmia Lake and Gavkhouni basins. Based on the graphical results, the estimated E_{pan} values of Shahid-Kazemi and Morcheh-Khort have maximum values of the median for Urmia Lake and Gavkhouni basins, respectively. Lighvan and Pol-Zamankhan stations have the lower estimated data variation coefficient for Dsa and Bsk climate type, respectively.

The special and unique structure of each model is the main factor in the difference in the performance of models. Because the basis of the GEP method is generation and chromosome production, this method has a high priority over the performance of other models. Also, the ANFIS method, which involves a combination of fuzzy rules and neuron structure, has high relative performance compared to the ANN method, which only includes the structure of neurons. By eliminating the data noise and the ability to establish very complex nonlinear relationships in its structure, wavelet analysis greatly enhances the performance of single models in all methods and climate types. Prioritization of model performance corresponds to different climatic conditions. Climatic conditions only affect the performance amount of the models but not their priority. However, the results of the three methods and their wavelet-hybrid could be acceptable for estimating the monthly E_{pan} variable in basins with Dsa and Bsk climatic types. On the other hand, in current research, applying the seasonal coefficient for both Dsa and Bsk climatic types is another factor that makes the results acceptable and strongly to estimate the monthly E_{pan}. It is necessary to mention that Dsa is a climate where there is at least 1 month colder than − 3 °C and summers are dry and hot. This kind of climate is usually at high elevations near locations that are warm temperate with dry, hot summers. Also, Bsk is a climate whose mean annual temperature is less than 18 °C and is too dry to support a forest, but not dry enough to be a desert, usually consisting of grassland plains.

The most important and practical advantage of the GEP and WGEP compared to other data-driven methods (ANFIS, ANN, WANFIS, and WANN) is in developing predictive equations to show the mathematical relationships between the model inputs and outputs. The development of mathematical relations

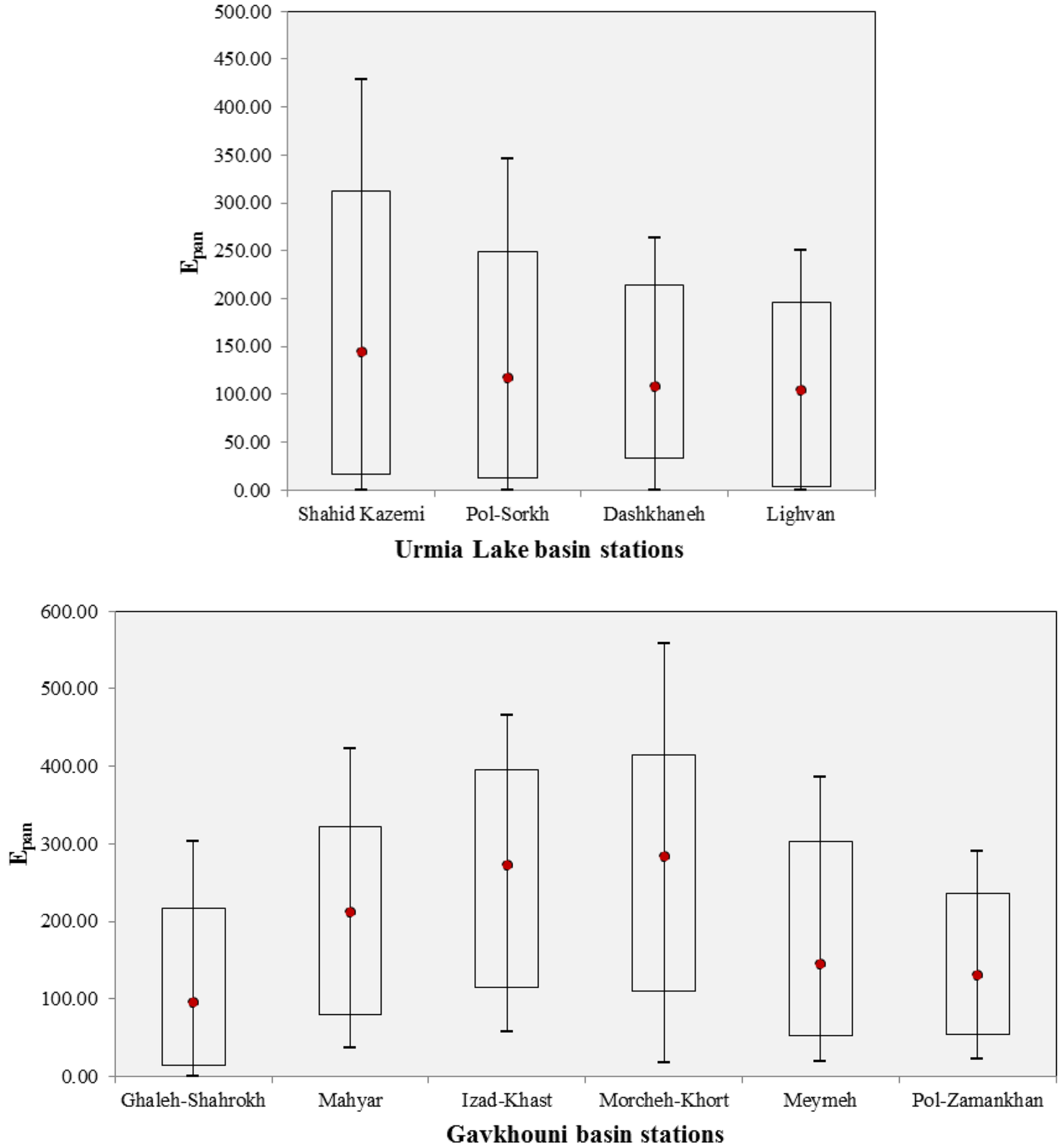


Fig. 15 Boxplot of the estimated monthly E_{pan} for the Urmia Lake and Gavkhouni basin and testing section

derived from the method of GEP and WGEP in various climates can provide a clear framework for determining the influencing factors on the rate of E_{pan} estimation in the region at different time and space scales. Also, these equations in the form of an affliction framework can be a good alternative to experimental mathematical relations and avoid wasting time and money. Table 5 lists the equations obtained with the optimal GEP and WGEP models for the monthly E_{pan} variables for both studied basins with various climatic conditions. The fitted equations can be applied at variable spatial and temporal scales, practically. Table 6

Table 6 The ranking of models results by climate type

Basin	Dsa (Urmia Lake)					
Stations	Shahid Kazemi	Pole-Sorkh	Dashkhaneh	Lighvan		
Ranking						
1	$WGEP_{11}$	$WGEP_{12}$	$WGEP_{13}$	$WGEP_{14}$		
2	$WANFIS_{11}$	$WANFIS_{12}$	$WANFIS_{13}$	GEP_{14}		
3	$WANN_{11}$	GEP_{12}	$WANN_{13}$	$WANFIS_{14}$		
4	GEP_{11}	$ANFIS_{12}$	GEP_{13}	$WANN_{14}$		
5	$ANFIS_{11}$	$WANN_{12}$	$ANFIS_{13}$	$ANFIS_{14}$		
6	ANN_{11}	ANN_{12}	ANN_{13}	ANN_{14}		
Basin	Bsk (Gavkhouni)					
Stations	Ghaleh-Shahrokh	Mahyar	Izad-Khast	Moorch-Khort	Meymeh	Pole-Zamankhan
Ranking						
1	$WGEP_{21}$	$WGEP_{22}$	$WGEP_{23}$	$WGEP_{24}$	$WGEP_{25}$	$WGEP_{26}$
2	$WANFIS_{21}$	$WANFIS_{22}$	GEP_{23}	$WANFIS_{24}$	GEP_{25}	$WANFIS_{26}$
3	GEP_{21}	GEP_{22}	$WANFIS_{23}$	GEP_{24}	$WANFIS_{25}$	GEP_{26}
4	$WANN_{21}$	$WANN_{22}$	$WANN_{23}$	$WANN_{24}$	$ANFIS_{25}$	$ANFIS_{26}$
5	$ANFIS_{21}$	$ANFIS_{22}$	$ANFIS_{23}$	$ANFIS_{24}$	$WANN_{25}$	$WANN_{26}$
6	ANN_{21}	ANN_{22}	ANN_{23}	ANN_{24}	ANN_{25}	ANN_{26}

summarizes the model's performance by ranking for each studied station of Dsa and Bsk climate classification, respectively. The results indicated that WGEP and ANN are the best and poorest models for all stations without affection for climate conditions in both basins. The selected models in this study were used to predict E_{pan} in two different climates.

Abghari et al. (2012) applied ANN and WANN models for estimating daily pan evaporation. Their results reflected the fact that the performance improvement of WANN compared to ANN models were varied from R: 3.88% to 9.60% and RMSE: 7.65% to 20.99% for the testing section. Qasem et al. (2019) applied wavelet support vector regression and wavelet artificial neural networks for estimating monthly pan evaporation at Tabriz and Antalya stations. Their results showed that performance improvement of wavelet hybrid models compared to single models was varied from R: − 1.12% to 0.10%, RMSE: − 23.63% to 5.5%, MAE: − 34.03% to 8.21%, and NSE: − 2.28% to 0.31% for Tabriz station and from R: − 0.53% to 0.76%), RMSE: − 9.2% to 4.61%, MAE: − 13.14% to 5.65%, and NSE: − 2.74% to 1.79% for Antalya station. Ghaemi et al. (2019) applied MARS, WMARS, M5, and WM5 methods for estimating monthly pan evaporation. Their results showed that the performance improvement of WM5 compared to M5 models was varied from R: − 18.35% to 1.61%, RMSE: − 57.60% to 23.42%, MAE: − 71.96% to 29.66%, and NSE: − 33.46% to 7.35 for validating section. Also, the performance improvement of WMARS compared to MARS models were varied from R: − 19.75% to 2.70%, RMSE: − 134.33% to 26.75%, MAE: − 126.80% to 18.04%, and NSE: − 53.61% to 9.38 for validating section.

Our results showed that the performance improvement of applied models (wavelet-hybrid models compared to single models) are ranged from R^2: 0.00% to 6.53%, RMSE: 2.77% to 34.79%, MAE: 0.02% to 29.61%, and NSE: 0.11% to 9.46% for all study stations in both climate categories. By comparing the performance improvement of different criteria, it can be achieved that the elimination of data noises in all cases has improved the performance of the models. In previous studies, which only presented a combination of models with wavelet theory without noise elimination, the performance improvement was either inverse or direct. Also, due to the uncertainty in hydrological phenomena and the need to remove them from modeling confirms the usability of the introduced new structure of modeling in current research. Besides, the improvement amount of various criteria is consistent with previous studies and showed the optimal performance of the models. Also, the special importance of the criteria, which is including the error of the models such as RMSE and MAE (Willmott & Matsuura, 2005; Chai and Draxler, 2014), and the need for a combined and overall evaluation of the used performance criteria, including R^2, RMSE, MAE, and NSE criteria, justifies the percentage of performance improvement of the wavelet-hybrid models compared to the single ones. These results indicate the applicability of wavelet-hybrid methods by eliminating the important factor of uncertainty in hydrological phenomena. On the other hand, the execution time of wavelet-hybrid models compared to single models has relative reductions that were varied from 1 to 5%. In addition to the selection of meteorological variables, the choice of seasonal coefficient variable as input variables can also improve the performance of wavelet-hybrid models. A comparison of the results of similar models in the two climates showed that due to the structure of the models, the type of climate did not make a significant difference in the model's results. The reason is that data-driven models with their unique structures could have the ability to predict in both types of regions. Elimination of noises, caused by human and natural factors, from the measured data can bring their values closer to the amount of real observations in nature, which in turn improves the performance of wavelet-hybrid models. The results of this study are consistent with the results of the findings by Pammar and Deka (2017), Araghi et al. (2020), and Malik et al. (2020).

Conclusion

In this study, ANN, ANFIS, and GEP models, as well as a hybrid of these methods with wavelet theory, were used to calculate the amount of monthly E_{pan} in basins with various climate categories. Four stations from the Urmia Lake basin and six stations from the Gavkhouni basin were selected based on the proper distribution of station at a 99% probability level using F-computational values to apply and assess the models. To increase the accuracy of the monthly E_{pan} estimation, noise elimination of climate data was performed and seasonal coefficients were used for each station with different climate conditions.

The performance accuracy of the different applied models in this study was evaluated with R^2, RMSE, MAE, and NSE statistics. The results of R^2 values for single and wavelet-hybrid data-driven models are close to 1, with the quality relations being: $R^2_{GEP} > R^2_{ANFIS} > R^2_{ANN}$ and $R^2_{WGEP} > R^2_{WANFIS} > R^2_{WANN}$ for all stations studied basins. Results showed that the GEP method has a high priority in terms of performance compared to the other models. In addition, performance of single models in all methods was greatly enhanced by eliminating the data noise and establishing complex nonlinear relationships in wavelet structure. Due to the mathematical relationship that can be used at different time and places, the WGEP model had a high priority over other models in terms of performance.

In the continuation of this research, it is suggested that researchers estimate E_{pan} by empirical formulas and compare the results of these models with data-driven methods. It is also recommended that E_{pan} be estimated on a daily scale and various climate categories to provide a strong framework of data-driven models for different climate conditions. This study can have a significant impact on the management, monitoring, and planning of water resource policymakers. Scientists can achieve more accurate results with fewer errors based on proper performance and high accuracy of estimation in AI methods than experimental methods. Also, the presented mathematical relationships in the GEP and WGEP methods show the important variables influencing the estimation of the E_{pan}. The organizations such as meteorological, water resources management, and environmental organizations can benefit from the results of current research (see Figs. 16, 17, 18, 19 for more detail of wavelet analysis decomposition and Table 7 for the values of correlation between used data).

Funding The Sari Agricultural Science and Natural Resources University (SANRU) financed this research (Grant No. 02–1399-15).

Data availability All of the required data have been presented in our article.

Declarations

Conflict of interest The authors declare no competing interests.

Appendix

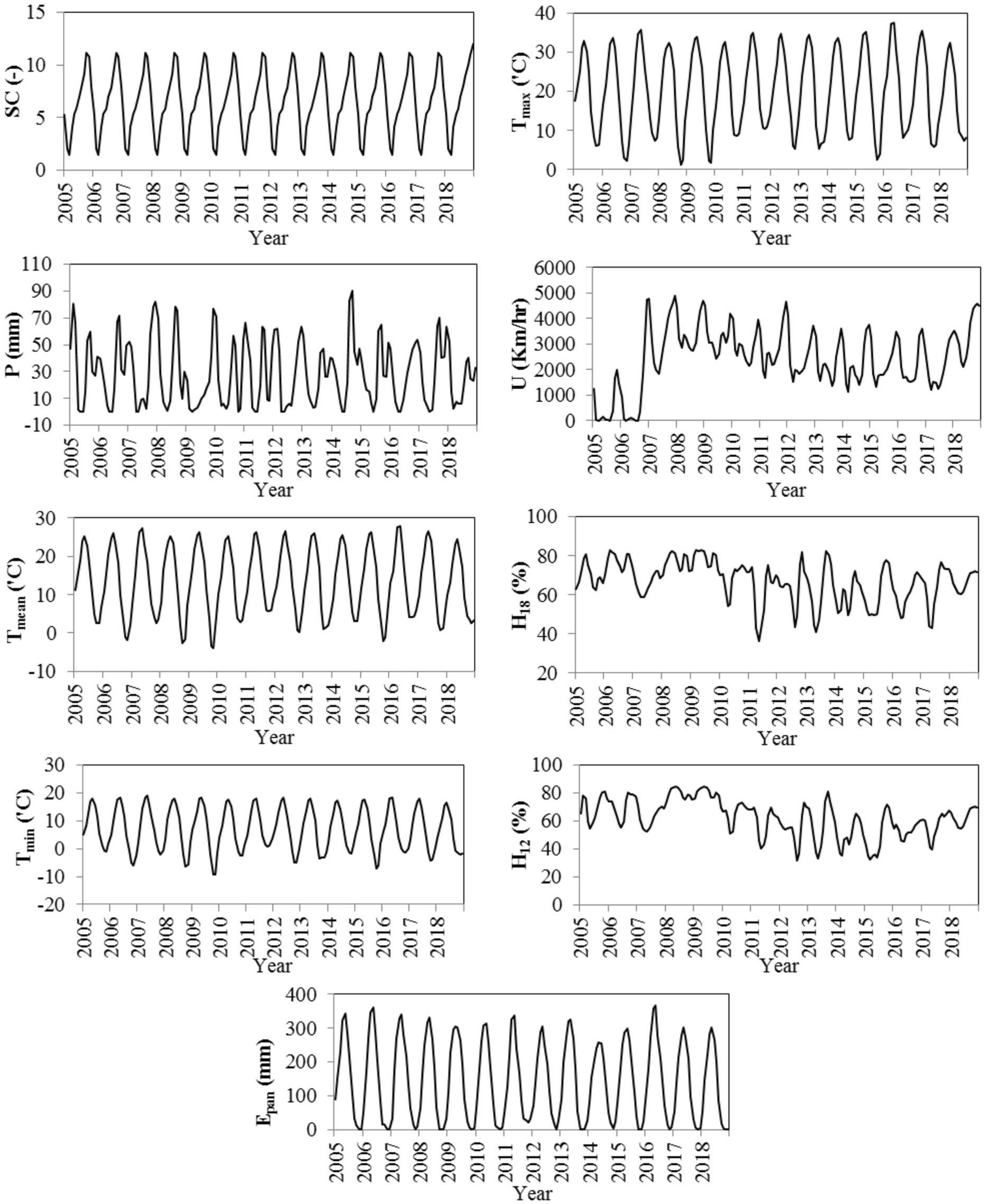


Fig. 16 A1 wavelet analysis decomposition for Pol-Sorkh station in Urmia Lake basin

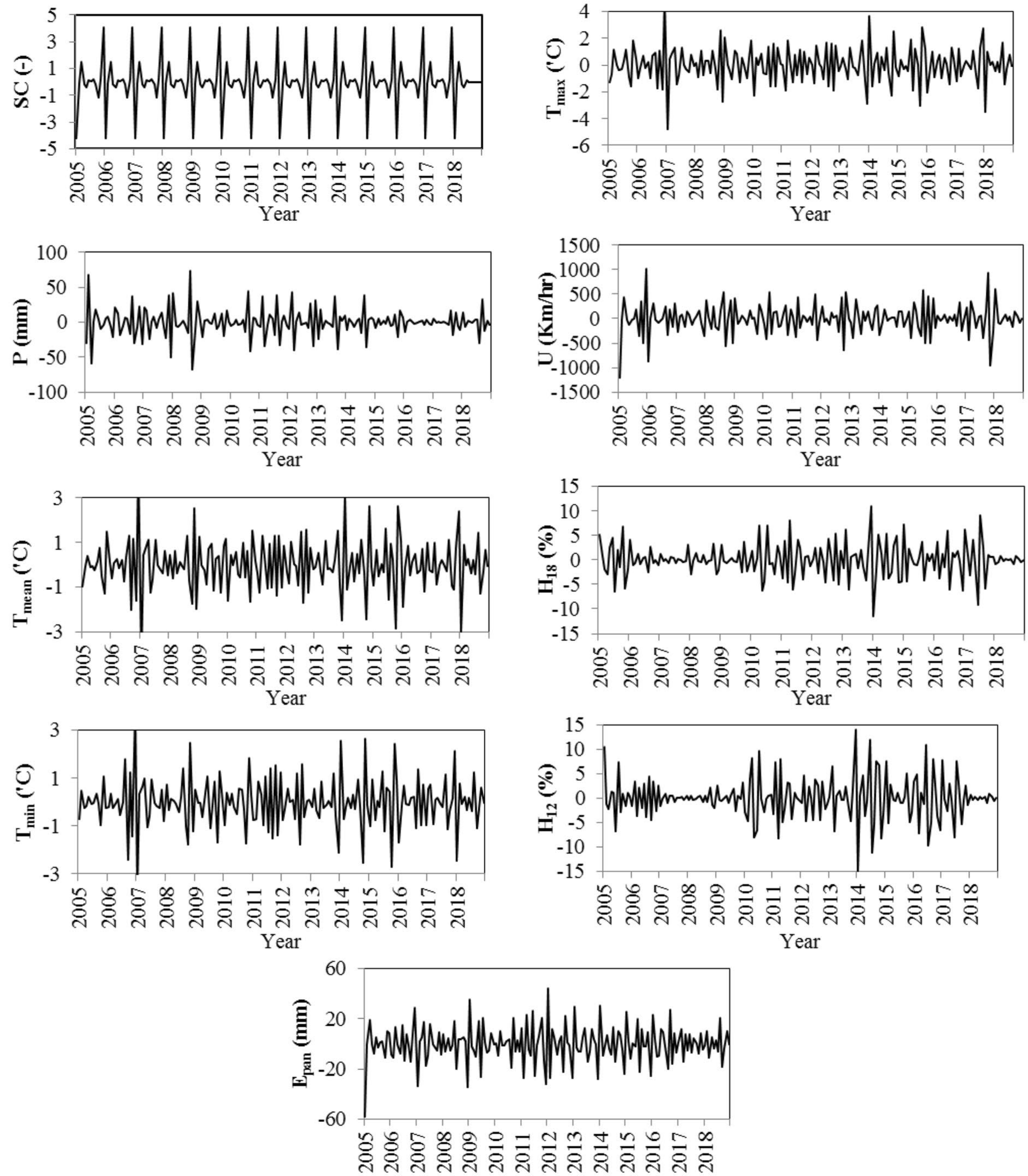


Fig. 17 D1 wavelet analysis decomposition for Pol-Sorkh station in Urmia Lake basin

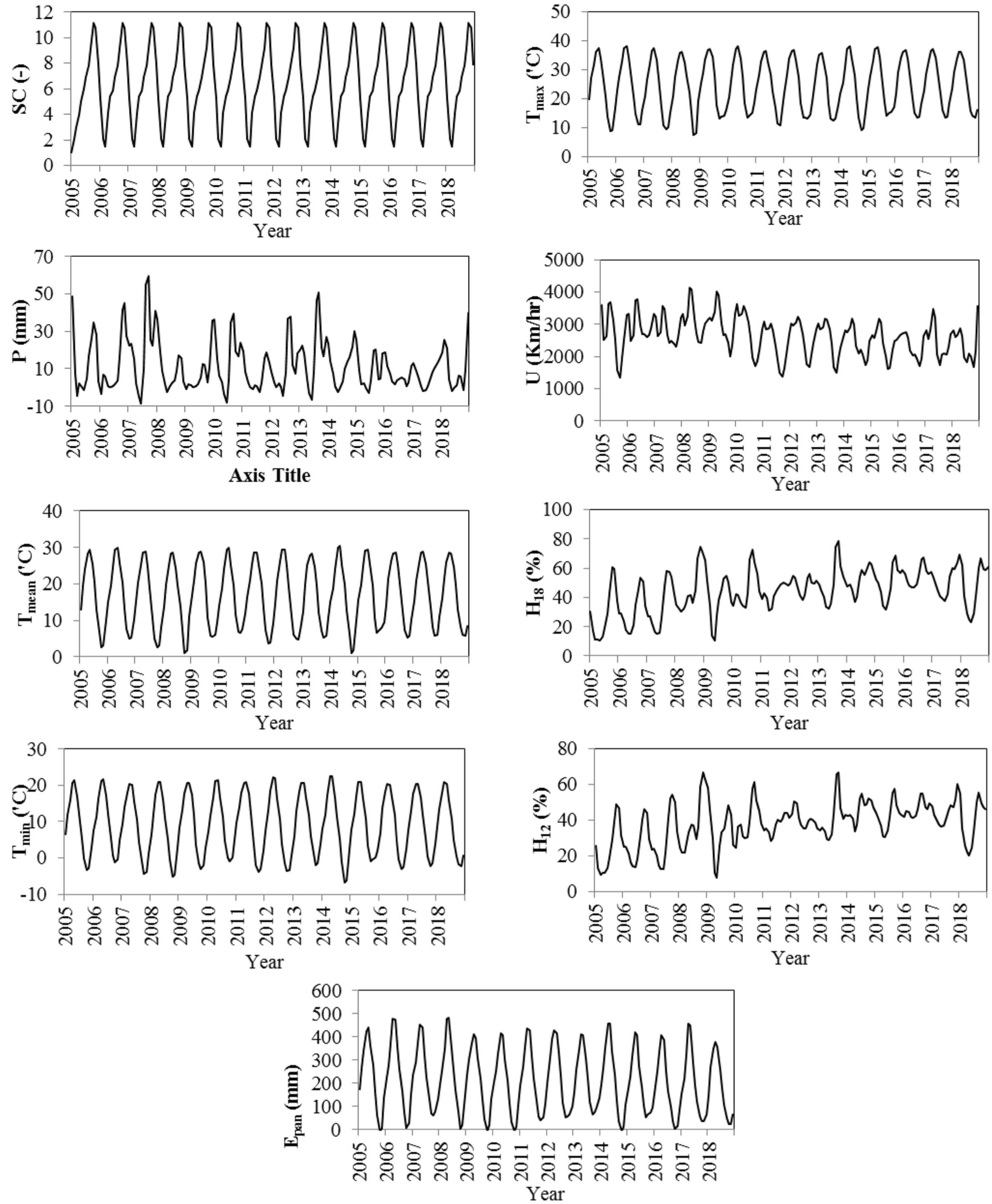


Fig. 18 A1 wavelet analysis decomposition for Mahyar station in Gavkhouni basin

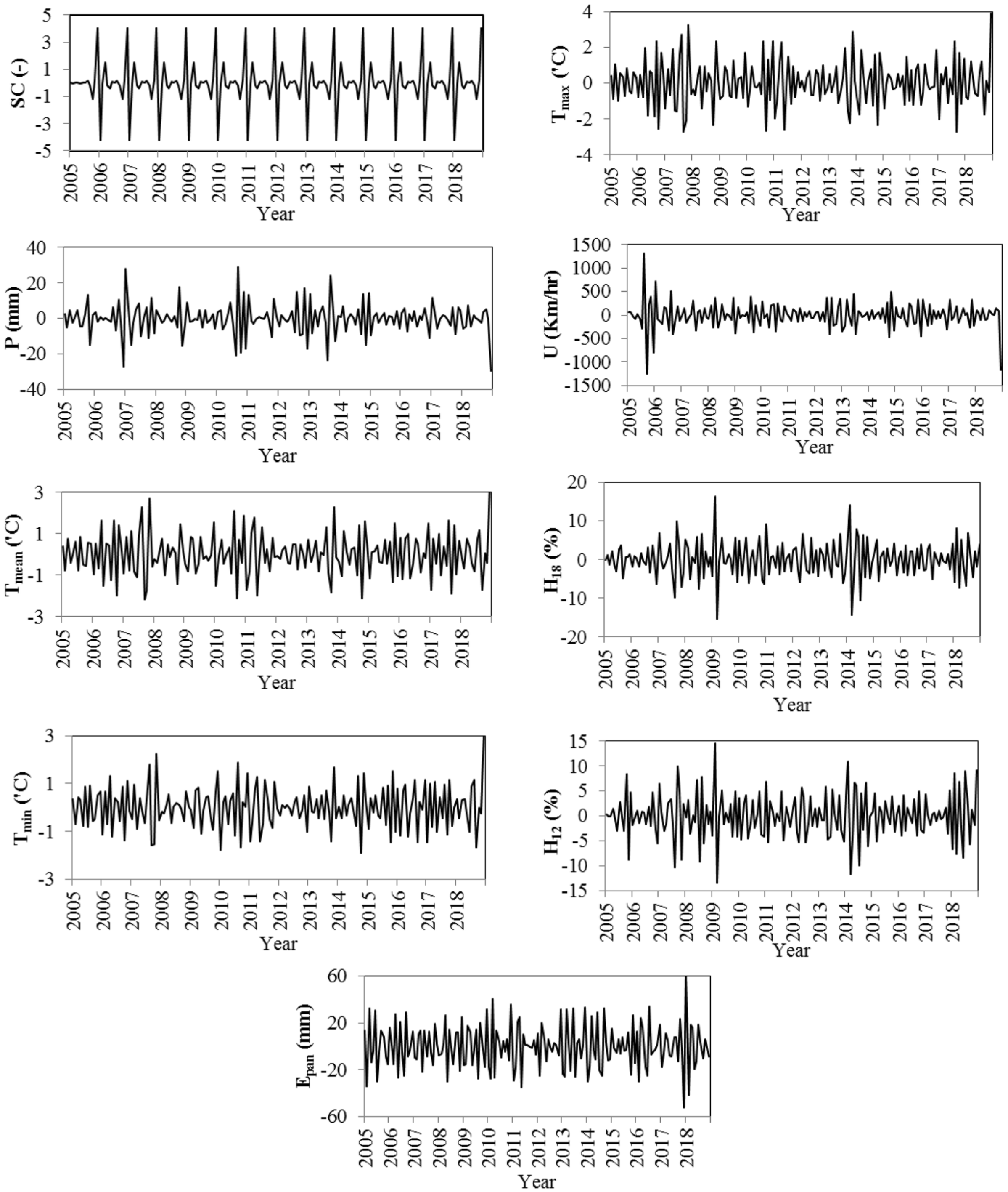


Fig. 19 D1 wavelet analysis decomposition for Mahyar station in Gavkhouni basin

Table 7 The absolute value of Pearson correlation coefficient between variables and E_{pan}

Basin	Urmia Lake				Gavkhouni					
Variables	Shahid Kazemi	Pol-Sorkh	Dashkhaneh	Lighvan	Ghaleh-Shahrokh	Mahyar	Izad-Khast	Morcheh-Khort	Meymeh	Pol-Zamankhan
P	0.572^{**}	0.553^{**}	0.260^{**}	0.271^{**}	0.610^{**}	0.459^{**}	0.504^{**}	0.341^{**}	0.621^{**}	0.506^{**}
T_{mean}	0.958^{**}	0.963^{**}	0.936^{**}	0.961^{**}	0.911^{**}	0.961^{**}	0.962^{**}	0.913^{**}	0.965^{**}	0.941^{**}
T_{min}	0.948^{**}	0.959^{**}	0.944^{**}	0.942^{**}	0.868^{**}	0.965^{**}	0.957^{**}	0.924^{**}	0.970^{**}	0.936^{**}
T_{max}	0.961^{**}	0.960^{**}	0.944^{**}	0.952^{**}	0.922^{**}	0.950^{**}	0.956^{**}	0.887^{**}	0.964^{**}	0.945^{**}
U	0.232^{**}	0.456^{**}	0.285^{**}	0.262^{**}	0.290^{**}	0.620^{**}	0.325^{**}	0.593^{**}	0.622^{**}	0.268^{**}
H_{18}	0.804^{**}	0.389^{**}	0.613^{**}	0.831^{**}	0.846^{**}	0.696^{**}	0.931^{**}	0.806^{**}	0.869^{**}	0.793^{**}
H_{12}	0.649^{**}	0.422^{**}	0.501^{**}	0.719^{**}	0.738^{**}	0.612^{**}	0.892^{**}	0.763^{**}	0.853^{**}	0.768^{**}

**Correlation is significant at the 0.01 level (2-tailed)

References

Abghari, H., Ahmadi, H., Besharat, S., & Rezaverdinejad, V. (2012). Prediction of daily pan evaporation using wavelet neural networks. *Water resources management, 26*(12), 3639–3652

Adnan, R. M., Heddam, S., Yaseen, Z. M., Shahid, S., Kisi, O., & Li, B. (2021). Prediction of potential evapotranspiration using temperature-based heuristic approaches. *Sustainability, 13*(1), 297

Adnan, R. M., Malik, A., Kumar, A., Parmar, K. S., & Kisi, O. (2019). Pan evaporation modeling by three different neurofuzzy intelligent systems using climatic inputs. *Arabian Journal of Geoscience, 12*, 606–614

Ahmadaali, Kh., Liaghat, A.M., Heydari, N., & Bozorg-Haddad, O. (2013). Application of artificial neural network and adaptive neural-based fuzzy inference system techniques in estimating of virtual water. *International Journal of Computer Applications* (0975 – 8887), 76 (6)

Araghi, A., Adamowski, J., & Martinez, C. J. (2020). Comparison of wavelet-based hybrid models for the estimation of daily reference evapotranspiration in different climates. *Journal of Water and Climate Change, 11*(1), 39–53

Ashrafzadeh, A., Ghorbani, M. A., Biazar, S. M., & Yaseen, Z. M. (2019). Evaporation process modelling over northern Iran: application of an integrative data-intelligence model with the krill herd optimization algorithm. *Hydrological Sciences Journal, 64*(15), 1843–1856

Azamathulla, H. M., Rathnayake, U., & Shatnawi, A. (2018). Gene expression programming and artificial neural network to estimate atmospheric temperature in Tabuk, Saudi Arabia. *Applied Water Science, 8*(6), 1–7

Burt, C. M., Mutziger, A. J., Allen, R. G., & Howell, T. A. (2005). Evaporation research: Review and interpretation. *Journal of Irrigation and Drainage Engineering, 131*(1), 37–58

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?–arguments against avoiding RMSE in the literature. *Geoscientific model development, 7*(3), 1247–1250

Chia, M. Y., Huang, Y. F., Koo, C. H., & Fung, K. F. (2020). Recent Advances in evapotranspiration estimation using artificial intelligence approaches with a focus on hybridization techniques—a review. *Agronomy, 10*(1), 101

Danandeh Mehr, A., Kahya, E., & Olyaie, E. (2013). Streamflow prediction using linear genetic programming in comparison with a neuro-wavelet technique. *Journal of Hydrology, 505*, 240–249

Eslamian, S. S., Gohari, S. A., Biabanaki, M., & Malekian, R. (2008). Estimation of monthly pan evaporation using artificial neural networks and support vector machines. *Journal of Applied Science, 8*(19), 3497–3502

Ghaemi, A., Rezaie-Balf, M., Adamowski, J., Kisi, O., & Quilty, J. (2019). On the applicability of maximum overlap discrete wavelet transform integrated with MARS and M5 model tree for monthly pan evaporation prediction. *Agricultural and Forest Meteorology, 278*, 107647

Ghavidel, S. Z. Z., & Montaseri, M. (2014). Application of different data-driven methods for the prediction of total dissolved solids in the Zarinehroud basin. *Stochastic*

Environmental Research and Risk Assessment, 28(8), 2101–2118

Guven, A., & Kisi, O. (2013). Monthly pan evaporation modeling using linear genetic programming. *Journal of Hydrology, 503*, 178–185

Holland, J.H. (1975). Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control and artificial intelligence. MIT Press, 1992. First Published by University of Michigan Press, USA

Huo, Z., Feng, Sh., Kang, Sh., Huang, G., Wanga, F., & Guo, P. (2012). Integrated neural networks for monthly river flow estimation in arid inland basin of Northwest China. *Journal of Hydrology, 420–421*, 159–170

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of Forecasting, 22*(4), 679–688

Jang, J.-S.R. (1993). ANFIS: adaptive network based fuzzy inference system. *IEEE Transactions on Systems, Man and Cybernetics., 23*(3), 665–683

Jing, W., Yaseen, Z. M., Shahid, S., Saggi, M. K., Tao, H., Kisi, O., & Chau, K. W. (2019). Implementation of evolutionary computing models for reference evapotranspiration modeling: short review, assessment and possible future research directions. *Engineering Applications of Computational Fluid Mechanics, 13*(1), 811–823

Keshtegar, B., Piri, J., & Kisi, O. (2016). A nonlinear mathematical modeling of daily pan evaporation based on conjugate gradient method. *Computers and Electronics in Agriculture, 127*, 120–130

Khaniya, B., Priyantha, H. G., Baduge, N., Azamathulla, H. M., & Rathnayake, U. (2020). Impact of climate variability on hydropower generation: a case study from Sri Lanka. *ISH Journal of Hydraulic Engineering, 26*(3), 301–309

Khatibi, R., Ghorbani, M. A., Naghshara, S., Aydin, H., & Karimi, V. (2020). Introducing a framework for 'inclusive multiple modelling' with critical views on modelling practices-Applications to modelling water levels of Caspian Sea and Lakes Urmia and Van. *Journal of Hydrology*, 124923

Kisi, Ö. (2006). Daily pan evaporation modelling using a neuro-fuzzy computing technique. *Journal of Hydrology, 329*(3–4), 636–646

Kisi, O., & Aliza mir, M. . (2018). Modelling reference evapotranspiration using a new wavelet conjunction heuristic method: Wavelet extreme learning machine vs wavelet neural networks. *Agricultural and Forest Meteorology, 263*, 41–48

Kisi, O., & Haddam, S. (2019). Evaporation modelling by heuristic regression approaches using only temperature data. *Hydrological Sciences Journal, 64*(6), 653–672

Kisi, O., Shiri, J., & Tomball, M. (2013). Modeling rainfall-runoff process using soft computing techniques. *Computers and Geosciences, 51*, 108–117

Kucuk, M., & Agiralioglu, N. (2006). Wavelet regression techniques for stream flow predictions. *Journal of Applied. Statistics., 33*(9), 943–960

Majidi, M., Alizadeh, A., Farid, A., & Vazifedoust, M. (2015). Estimating evaporation from lakes and reservoirs under limited data condition in a semi-arid region. *Water Resources Management, 29*(10), 3711–3733

Malik, A., & Kumar, A. (2015). Pan evaporation simulation based on daily meteorological data using soft computing techniques and multiple linear regression. *Water Resources Management, 29*(6), 1859–1872

Malik, A., Kumar, A., Kim, S., Kashani, M. H., Karimi, V., Sharafati, A., & Chau, K. W. (2020). Modeling monthly pan evaporation process over the Indian central Himalayas: application of multiple learning artificial intelligence model. *Engineering Applications of Computational Fluid Mechanics, 14*(1), 323–338

Malik, A., Kumar, A., & Kisi, O. (2017). Monthly pan-evaporation estimation in Indian central Himalayas using different heuristic approaches and climate-based models. *Computers and Electronics in Agriculture, 143*, 302–313

Mamdani, E. H., & Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies, 7*, 1–13

McColl, K. A. (2020). Practical and theoretical benefits of an alternative to the Penman-Monteith evapotranspiration equation. *Water Resources Research*, *56*(6), e2020WR027106

Moghaddamnia, A., Gousheh, M. G., Piri, J., Amin, S., & Han, D. (2009). Evaporation estimation using artificial neural networks and adaptive neuro-fuzzy inference system techniques. *Advances in Water Resources, 32*(1), 88–97

Montaseri, M., Ghavidel, S. Z. Z., & Sanikhani, H. (2018). Water quality variations in different climates of Iran: toward modeling total dissolved solid using soft computing techniques. *Stochastic Environmental Research and Risk Assessment, 32*(8), 2253–2273

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

Pammar, L., & Deka, P. C. (2017). Daily pan evaporation modeling in climatically contrasting zones with hybridization of wavelet transform and support vector machines. *Paddy and Water Environment, 15*(4), 711–722

Perera, A. N. U. S. H. K. A., Azamathulla, H., & Rathnayake, U. P. A. K. A. (2020a). Comparison of different artificial neural network (ANN) training algorithms to predict atmospheric temperature in Tabuk, Saudi Arabia. *MAUSAM: Quarterly Journal of Meteorology, Hydrology & Geophysics*, *71*(2), 551–560

Perera, A., Mudannayake, S. D., Azamathulla, H. M., & Rathnayake, U. (2020b). Recent climatic trends in Trinidad and Tobago, West Indies. *Asia-Pacific Journal of Science and Technology*, *25*(02)

Popoola, A.O. (2007). Fuzzy-wavelet method for time series analysis. University of Surrey. Department of Computing School of Electronics and Physical Sciences University of Surrey Guildford, Surrey GU2 7XH, UK

Qasem, S. N., Samadianfard, S., Kheshtgar, S., Jarhan, S., Kisi, O., Shamshirband, S., & Chau, K. W. (2019). Modeling monthly pan evaporation using wavelet support vector regression and wavelet artificial neural networks in arid and humid climates. *Engineering Applications of Computational Fluid Mechanics, 13*(1), 177–187

Qutbudin, I., Shiru, M. S., Sharafati, A., Ahmed, K., Al-Ansari, N., Yaseen, Z. M., & Wang, X. (2019). Seasonal drought pattern changes due to climate variability: Case study in Afghanistan. *Water, 11*(5), 1096

Or, D., & Lehmann, P. (2019). Surface evaporative capacitance: How soil type and rainfall characteristics affect global-scale surface evaporation. *Water Resources Research, 55*(1), 519–539

Ozger, M. (2010). Significant wave height forecasting using wavelet fuzzy logic approach. *Ocean Engineering., 37*, 1443–1451

Partal, T., & Kisi, O. (2007). Wavelet and neuro-fuzzy conjunction model for precipitation forecasting. *Journal of Hydrology, 342*, 199–212

Rezaie-Balf, M., Kisi, O., & Chua, L. H. (2019). Application of ensemble empirical mode decomposition based on machine learning methodologies in forecasting monthly pan evaporation. *Hydrology Research, 50*(2), 498–516

Sanikhani, H., Kisi, O., Kiafar, H., & Ghavidel, S. Z. Z. (2015). Comparison of different data-driven approaches for modeling lake level fluctuations: the case of Manyas and Tuz Lakes (Turkey). *Water Resources Management, 29*(5), 1557–1574

Sanikhani, H., Kisi, O., Nikpour, M. R., & Dinpashoh, Y. (2012). Estimation of daily pan evaporation using two different adaptive neuro-fuzzy computing techniques. *Water Resources Management, 26*(15), 4347–4365

Sebbar, A., Heddam, S., & Djemili, L. (2019). Predicting daily pan evaporation (E pan) from dam reservoirs in the Mediterranean Regions of Algeria: OPELM vs OSELM. *Environmental Processes, 6*(1), 309–319

Seo, Y., Kim, S., Kisi, O., & Singh, V. P. (2015). Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. *Journal of Hydrology, 520*, 224–243

Shiri, J. (2019). Evaluation of a neuro-fuzzy technique in estimating pan evaporation values in low-altitude locations. *Meteorological Applications, 26*(2), 204–212

Shirsath, P. B., & Singh, A. K. (2010). A comparative study of daily pan evaporation estimation using ANN, regression and climate based models. *Water Resources Management, 24*(8), 1571–1581

Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control, *IEEE Transactions on Systems, Man and Cybernetics*. SMC-15 (1), 116–132

Tan, Q. F., Lei, X. H., Wang, X., Wang, H., Wen, X., Ji, Y., & Kang, A. Q. (2018). An adaptive middle and long-term runoff forecast model using EEMD-ANN hybrid approach. *Journal of Hydrology, 567*, 767–780

Tao, H., Diop, L., Bodian, A., Djaman, K., Ndiaye, P. M., & Yaseen, Z. M. (2018). Reference evapotranspiration prediction using hybridized fuzzy model with firefly algorithm: Regional case study in Burkina Faso. *Agricultural water management, 208*, 140–151

Tezel, G., & Buyukyildiz, M. (2016). Monthly evaporation forecasting using artificial neural networks and support vector machines. *Theoretical and applied climatology, 124*(1–2), 69–80

Trajkovic, S., & Kolakovic, S. (2010). Comparison of simplified pan-based equations for estimating reference evapotranspiration. *Journal of Irrigation and Drainage Engineering, 136*(2), 137–140

Wang, L., Kisi, O., Hu, B., Bilal, M., Zounemat-Kermani, M., & Li, H. (2017). Evaporation modelling using different machine learning techniques. *International Journal of Climatology, 37*, 1076–1092

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research, 30*(1), 79–82

Yaseen, Z. M., Al-Juboori, A. M., Beyaztas, U., Al-Ansari, N., Chau, K. W., Qi, C., & Shahid, S. (2020). Prediction of evaporation in arid and semi-arid regions: a comparative study using different machine learning models. *Engineering applications of computational fluid mechanics, 14*(1), 70–89

Yaseen, Z. M., El-Shafie, A., Jaafar, O., Afan, H. A., & Sayl, K. N. (2015). Artificial intelligence-based models for streamflow forecasting: 2000–2015. *Journal of Hydrology, 530*, 829–844

ZamanZad-Ghavidel, S., Bozorg-Haddad, O., & Goharian, E. (2020). *Sustainability assessment of water resource systems using a novel hydro-socio-economic index (HSEI).* (pp. 1–48). Environment.

Zareie, S., & Fakherifard, A. (2014). Assessment of temporal relationship between meteorological and hydrological droughts by GP and ANFIS models in Sofi-Chi basin. *Water Resources Engineering Journal, 7*(21), 37–50 (In Persian).

Zounemat-Kermani, M., Kisi, O., Piri, J., & Mahdavi-Meymand, A. (2019). Assessment of Artificial Intelligence-Based Models and Metaheuristic Algorithms in Modeling Evaporation. *Journal of Hydrologic Engineering, 24*(10), 04019033