Optimization-Based Prediction Uncertainty Qualification of Climatic Parameters

VAHID NOURANI,[a,b,c] MINA SAYYAH-FARD,[b] SAMEH A. KANTOUSH,[a] KHAGENDRA P. BHARAMBE,[a] TETSUYA SUMI,[a] AND MOHAMED SABER[a]

[a] *Water Resources Research Center, Disaster Prevention Research Institute, Kyoto University, Uji, Japan*
[b] *Center of Excellence in Hydroinformatics, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran*
[c] *Faculty of Civil and Environmental Engineering, Near East University, Lefkosa, Turkey*

(Manuscript received 10 March 2023, in final form 12 July 2023, accepted 17 July 2023)

ABSTRACT: Point predictions of hydroclimatic processes through nonlinear modeling tools are associated with uncertainty. The main goal of this research was to construct prediction intervals (PIs) for nonlinear artificial neural network (ANN)-based models of evaporation and the standardized precipitation index (SPI). These are two critical indicators for climate for four stations in Iran (i.e., Tabriz, Urmia, Ardabil, and Ahvaz) to qualify their predicted uncertainty values (UVs). We used classical techniques of bootstrap (BS), mean variance estimation (MVE), and Delta, as well as an optimization-based method of lower upper bound estimation (LUBE), to construct and compare the PIs. The wavelet-based denoising method was also adopted to denoise input data, enhancing the modeling performance. The obtained results indicate the ability of the BS and LUBE methods to estimate the uncertainty bound. The Delta method mostly failed to find the desired coverage due to its narrow PIs. On the other hand, the MVE method, due to its wide bound, did not convey valuable information about uncertainty. According to the obtained results, denoising the input vector could enhance the PI quality in the modeling of the SPI by up to 76%. It was more prominent than reducing the UV for evaporation models, which was observed the most at the Ardabil station, up to 30%. The inherently more random nature of drought than the evaporation process was interpreted as the cause of this reaction. From the results, Urmia station seems the riskiest regarding drought ventures.

SIGNIFICANCE STATEMENT: The point predictions of evaporation and precipitation (in the form of SPI) are subject to uncertainty. The best way is to provide an area with the highest contingency as a prediction interval. The reduction in the width of such an interval leads to increased confidence in explaining and predicting these processes. We investigated different methods and found that by utilizing the optimization-based method for denoised inputs, uncertainty values of the output were conveyed better. Additionally, we concluded that the more random the process, the greater its uncertainty. A primary sense of the drought risk was made from the uncertainty perspective.

KEYWORDS: Optimization; Uncertainty; Climate prediction; Artificial intelligence

1. Introduction

Most hydroclimatological processes, such as evaporation and precipitation, two critical components in forming the spatial pattern of water resources, are randomly included in nature so that accurate prediction of how they will change in the future is challenging and full of uncertainty. Although quantifying the uncertainty associated with the point predictions of these processes is essential, few studies have addressed these issues. Such phenomena are nonlinear and complex. The flexibility of artificial neural networks (ANNs) as black box models for addressing complex problems makes it possible to capture the inherent nonlinearities of such processes without prior knowledge of their physics. Although many studies have already highlighted the ability of ANNs to make good point predictions of hydrologic processes (Tanty and Desmukh 2015; Chen et al. 2020), their output is associated with uncertainty. The nonoptimal ANN structure/parameters and noise of the observed data are the sources of the ANN output uncertainty values (UV).

Considering the uncertainty in the input as the most significant source of uncertainty for data-driven models, in this study, we used prediction intervals (PIs) to quantify the uncertainty in the output of ANN models for evaporation and the standard precipitation index (SPI). In the literature, PI construction techniques have been divided into two categories based on ANN models (Kabir et al. 2018). The first category includes multistep techniques such as the mean variance estimation (MVE) provided by Nix and Weigend (1994), Delta (Hwang and Ding 1997; de Veaux et al. 1998), and bootstrap (BS) proposed by Efron (1979) as classical methods. In addition to the problematic implementation, applying these methods requires basic assumptions such as homogeneous and customarily distributed noise for the Delta and MVE methods (Khosravi et al. 2011b; Kabir et al. 2018). The Hessian matrix calculation is time consuming in the Delta method. Additionally, the BS method has a high computational cost for large datasets (Khosravi et al. 2011b). The second category was recently developed based on optimization techniques, including the PI's direct construction (Khosravi et al. 2011a). The basic

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JHM-D-23-0043.s1.

Corresponding author: Vahid Nourani, nourani.vahid.63t@st.kyoto-u.ac.jp; vnourani@yahoo.com

DOI: 10.1175/JHM-D-23-0043.1

technique in this group is the lower upper bound estimation (LUBE) method; variations of the original version of LUBE were developed by modifying its cost function (e.g., Quan et al. 2014; Zhang et al. 2015). LUBE-based methods have some advantages over alternative approaches because the PIs are generated through an optimization problem similar to the point prediction method without any assumption (Kabir et al. 2018). Despite the importance of PIs for hydroclimatological studies (e.g., Srivastav et al. 2007; Kasiviswanathan et al. 2013; Kumar et al. 2015; Taormina and Chau 2015; Nourani et al. 2019, 2020) and in some other engineering fields (e.g., Lian et al. 2016; Wang et al. 2018), the use of LUBE for water issues has been limited. Furthermore, no studies have used Delta and MVE methods to estimate PIs for hydroclimatological models; therefore, their ability to quantify UV for water-dependent problems is unknown.

Although PI estimation is essential for water-dependent planning, their improvement is challenging in creating more confidence and less operational risk. One improvement step to achieve this goal is reducing the uncertainty in the collected data and information, which forms the core of our comprehension of any phenomenon. The present research sought an appropriate data preprocessing method to investigate the possibility of reducing output uncertainty in modeling evaporation and meteorological drought (via SPI) processes. Naturally observed time series are usually contaminated by noise (which appears as high-frequency component of the input vector), which can lead to deviations in the learning process of ANN models since, as a nonlinear model, ANN can remarkably magnify noise, even small values, when going ahead in time. There are various methods for denoising, but due to the multiresolution (multiscale) essence of hydroclimatological processes, the use of wavelet-based denoising could lead to successful simulations (see Campisi-Pinto et al. 2012; Guo et al. 2011; Nourani et al. 2014). Additionally, redundant input space variables can pull down models' performance in terms of increasing calculation mass, model size, and induction of unnecessary/misleading patterns. To eliminate redundant inputs, it is necessary to interpret the internal relations of the ANN model in terms of output sensitivity magnitude per input. One alternative method for this purpose is to use a derivation-based technique named the partial derivatives (PaD) method, which is a robust mathematical technique for sensitivity analysis (see, e.g., Gedeon 1997; Gevrey et al. 2003, 2006; Azzahari et al. 2016). Since the ANN model is a kind of regression based model, the PaD technique tries to estimate the partial derivatives of the output on the input space and the magnitude of the produced PaD values (Gevrey et al. 2003; Lu et al. 2001).

Meteorological drought modeling and qualification of its uncertainty, the topic of this study, is an essential aspect in drought risk management. To monitor meteorological drought, precipitation anomalies can be considered (Pazhanivelan et al. 2023; Sahbeni et al. 2023; Aryal et al. 2022). However, drought modeling using precipitation anomalies (rainfall deviation from its long-term mean) could be limited by its intrinsic reliance on the mean values; thus, it cannot be applied uniformly to different areas with different amounts of mean rainfall values. Therefore, rainfall deviations across space and time must be interpreted with caution. Instead of precipitation, we used the SPI, which is universally used to monitor meteorological drought and has been recommended by World Meteorological Organization (WMO 2012) as well as some other studies (e.g., Koutroulis et al. 2011; Kumar et al. 2009; Pande et al. 2022; Pathak and Dodamani 2020; Zhou et al. 2023). On the other hand, because evaporation is a crucial climatic factor for the hot areas when investigating SPI, regional evaporation should also be surveyed, and it helps to have a separate view of the climate-forcing elements. Therefore, in this study, we focused on evaporation and precipitation (from the SPI perspective) as particular meteorological elements. Nevertheless, one can use both of them in a unique framework, such as the standardized precipitation evapotranspiration index (SPEI) as a multivariate index to monitor drought (Tan et al. 2023; Kamruzzaman et al. 2022). In summary, this study aims to address three questions:

1) Which process under investigation (evaporation or SPI which include different magnitude of high-frequency components as disturbance or randomness) does include more uncertainty and why?
2) Which methods have performed better in estimating PIs?
3) What was the effect of input data denoising (removing high-frequency components) on the PIs?

All abbreviations and symbols used in this study (text and equations) are documented in appendixes A and B, respectively.

2. Materials and methods

a. Study site and data

In this study, the evaporation process and SPI were modeled at four climatic stations in Iran. Because droughts have a high probability of occurrence in the northwestern provinces of Iran, three stations named Tabriz, Ardabil, and Urmia were considered for the study, which have similar climates. In addition, the study also considered the Ahvaz station in a hot and humid climate region of Iran, which has particular conditions for evaporation and drought. Figure 1 shows the locations of the selected stations and their climatic classifications according to the Köppen criterion. Additionally, some geographic and climate aspects of the stations are tabulated in Table 1. From Fig. 1, the climate regions of Tabriz, Urmia, and Ardabil cities are cold and continental or Mediterranean, while the climate of Ahvaz city is hot and desert.

The daily gauge data (synoptic stations) used in this study were collected over the available timespans (Table 2) and included the air temperature [T (°C)], pan evaporation [E_t (mm day^{-1})], solar radiation [R_s (J cm^{-2} day^{-1})], air pressure [P (hPa)], relative humidity [R_h (%)], and wind speed [U (m s^{-1})], which were used to model E_t and estimate its PIs. Similarly, observed precipitation data [Pr (mm)] over the available period (Table 2) were used on a monthly scale to model the SPI and construct its PIs for these stations. All datasets used in this study were provided by the Iran Meteorological Organization (https://www.irimo.ir/index.php?newlang=eng). The statistical characteristics and timespans of each data variable are tabulated in Table 2. For any station, all datasets were allocated to 70% for the training group and 30% for the test group.

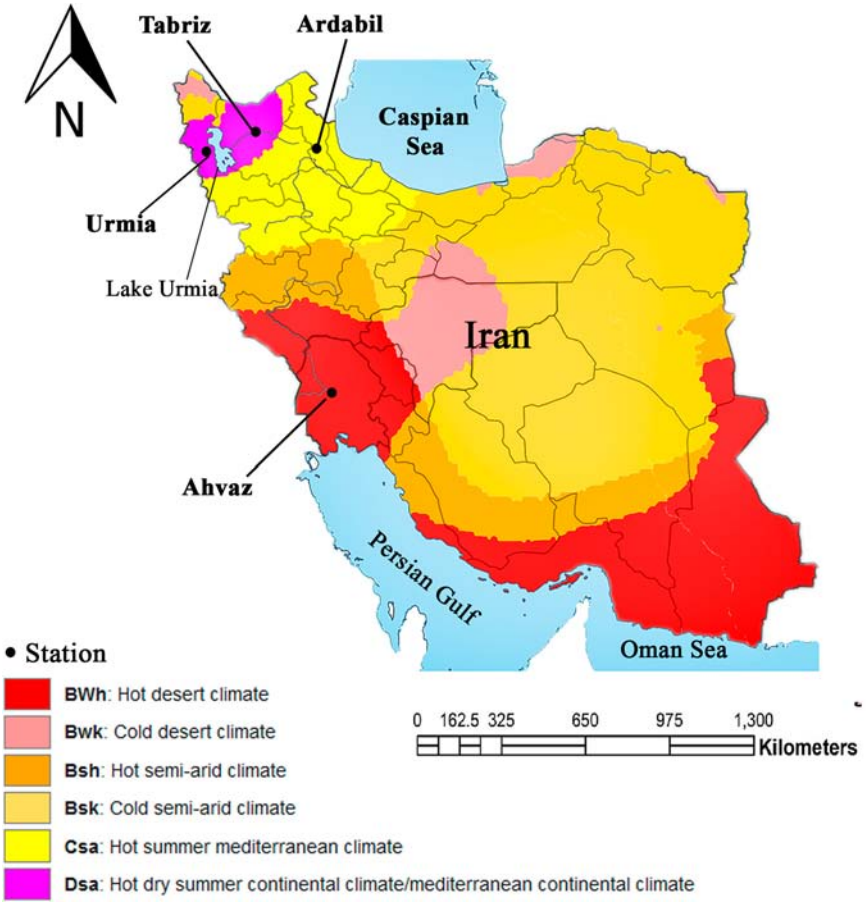

FIG. 1. Location of study stations along with their Koppen climate classification.

Despite the high T, the Ahvaz station experienced less E_t and Pr in the summer season than the other stations. Therefore, the R_h agent was dominant at this station during summer, leading to a reverse rate of E_t (and complex E_t pattern), which was also pointed out by previous studies (e.g., Nasrollahi et al. 2021). However, it seems that water stress is more prominent for the Urmia and Tabriz stations with little Pr but high E_t in the summer season because the propagation time of meteorological drought, as the most important aspect of drought, into hydrological and agricultural droughts can be shorter in the summer than in the cold seasons of the year (Li et al. 2021). The Ardabil station has the highest Pr in the warm period rather than the other stations, but its E_t is not high. Similar to the Ahvaz station, the R_h agent at this station is a crucial climatic factor.

To monitor medium- and long-term meteorological drought, the SPI values were used as 9-month SPI (SPI-9) and 12-, 24-, and 48-month SPI (SPI-12, SPI-24, and SPI-48). The SPI can be calculated over different rainfall accumulation periods. The different scales show the different potential importance in water resource management, such as SPIs for shorter time scales (e.g., from SPI-1 to SPI-3) are the indicators of immediate impacts such as reduced soil moisture; SPIs for medium time scales (e.g., from SPI-6 to SPI-9) are the indicators for reduced streamflow and reservoir storage, whereas the SPIs for longer time scales (e.g., from SPI-12 to SPI-48) are the indicators for reduced reservoir and groundwater recharge (McKee et al. 1993; Komuscu 1999). To calculate the SPIs, the Drought Indices Package (DIP) was used with monthly Pr data, and section S1 in the online

TABLE 1. Details about geographic locations and climate of the study area. "A A." stands for average annual.

	Geographic location			Climate	
Station	Longitude (°N)	Latitude (°W)	Height (m) above sea level	A A. Pr (mm)	A A. E_t (mm)
Tabriz	46.26	38.06	1328 to 2361	284.34	2248
Urmia	44.58	37.34	1332	338.34	1529
Ardabil	48.17	38.15	1351	287.38	1471
Ahvaz	48.67	31.31	16	214.10	3190

TABLE 2. Statistical representatives of the datasets used at all relevant synoptic stations. CV is the coefficient of variation.

	Timespan	Max	Min	Avg	CV (%)	Timespan	Max	Min	Avg	CV (%)
		Ahvaz					Tabriz			
E_t (mm day^{-1})	1998–2019	25.80	1.00	8.74	65	1992–2018	22.20	0.10	6.16	88
T (°C)	1998–2019	45.00	5.00	26.63	36	1992–2018	35.10	−15.00	13.34	78
R_h (%)	1998–2019	99.25	6.93	40.77	50	1992–2018	96.87	10.50	50.57	35
U (m s^{-1})	1998–2019	6.88	0.00	2.30	58	1992–2018	9.63	0.27	3.39	48
R_s (J cm^{-2} day^{-1})	1998–2019	3271.00	31.00	1856.90	37	1992–2018	3604.0	43.00	1698.20	47
P (hPa)	1998–2019	1024.89	989.14	1006.80	0.84	1992–2018	880.54	848.33	864.40	0.50
Pr (mm)	1957–2020	153.11	0.00	17.84	160	1951–2020	128.43	0.00	23.70	98
		Urmia					Ardabil			
E_t (mm day^{-1})	1992–2004, 2007–18	14.10	0.00	4.19	81	2006–18	14.70	0.00	4.03	79
T (°C)	1992–2004, 2007–18	30.00	−13.00	12.04	79	2006–18	26.70	−19.50	10.27	81
R_h (%)	1992–2004, 2007–18	100.00	15.62	57.87	27	2006–18	100.00	24.63	72.79	22
U (m s^{-1})	1992–2004, 2007–18	8.63	0.10	2.19	51	2006–18	14.50	0.15	3.69	69
R_s (J cm^{-2} day^{-1})	1992–2004, 2007–18	3540.00	16.00	1778.5	45	2006–18	3721.0	102.00	1661.00	45
P (hPa)	1992–2004, 2007–18	882.00	851.24	867.18	0.51	2006–18	882.00	850.50	866.26	0.49
Pr (mm)	1951–2020	138.88	0.00	28.20	107	1977–2020	122.9	0.00	23.95	92

supplemental material refers to the results. Dryness/wetness categories based on SPI intervals are shown in Table 3, outstanding the extreme states.

b. Proposed methodology

The proposed study procedure is illustrated in Fig. 2. Accordingly, dominant inputs were first selected and subsequently denoised by PaD and wavelet techniques, respectively. Then, suitable models were developed based on the ANNs for both raw and denoised inputs to estimate E_t and SPI. In the following, PIs were estimated by applying different methods and then compared. The components of the applied methodology are described briefly as follows.

1) POINT PREDICTIONS BY ANN

A neural network is a set of neurons that capture inputs and connect with the information of other nodes. In this "artificial brain," there are three or multiple layers (a deep learning network) with the building blocks of neurons where each node is connected to the subsequent layer nodes. Widely used algorithm, known as backpropagation (BP), trains the network to recognize correct patterns. In BPNN, training is based on the repeated updating of the network's parameters through the negative gradient of a mean squared error (MSE) cost function. The BP algorithm involves two passes through the ANN's layers: forward and reverse training. The difference between the desired and actual output values generates the error signal, which is backpropagated to the lower layers to update the BPNN parameters. The learning rate and momentum factor are used for controlling the weight adjustment along the steepest descent direction and for dampening oscillations (Zweiri et al. 2003; Singh et al. 2022). A neural network (NN) acts as a regression analysis that predicts future events based on past states (training dataset). To prevent overfitting of the ANN model on the training dataset and increase its generalization ability, in this study, a weight decay cost function with the regularization parameter λ and estimated weight vector $\hat{\mathbf{w}}$ was used according to Eq. (1). In this way, the larger coefficients of the input parameters are controlled by a "penalty" term, which subsequently leads to the limitation of the model's variance. To enhance the ANN generalization ability, in this study, λ in Eq. (1) was set to 0.9 (through trial and error):

$$\text{WDCF} = \text{MSE} + \lambda\, \hat{\mathbf{w}}^{\mathrm{T}}\, \hat{\mathbf{w}}. \tag{1}$$

In this study, the appropriate number of ANN training epochs was determined through trial and error; after this number of epochs, the models no longer improved. The statistical indices of root-mean-square error (RMSE), as well as the determination coefficient (DC) or Nash–Sutcliffe criterion, are usually sufficient to determine the model's performance for point predictions (Nourani et al. 2021) and were used in this paper to select the proper ANN structure. RMSE and DC belong to ranges of $[0, +\infty)$ and $(-\infty, 1]$ with desirable values near 0 and 1, respectively.

TABLE 3. Dryness/wetness categories based on SPI intervals (McKee et al. 1993).

Category	SPI intervals
Extremely wet (EW)	≥2.0
Severely wet (SW)	1.50 to 1.99
Moderate wet (MW)	1.00 to 1.49
Near normal	0.99 to −0.99
Moderately dry (MD)	−1.00 to −1.49
Severely dry (SD)	−1.50 to −1.99
Extremely dry (ED)	≤−2.0

2) DATA PREPROCESSING

The data preprocessing step is usually applied to the dataset before training the model and is desirable for saving training time and improving overall model performance.

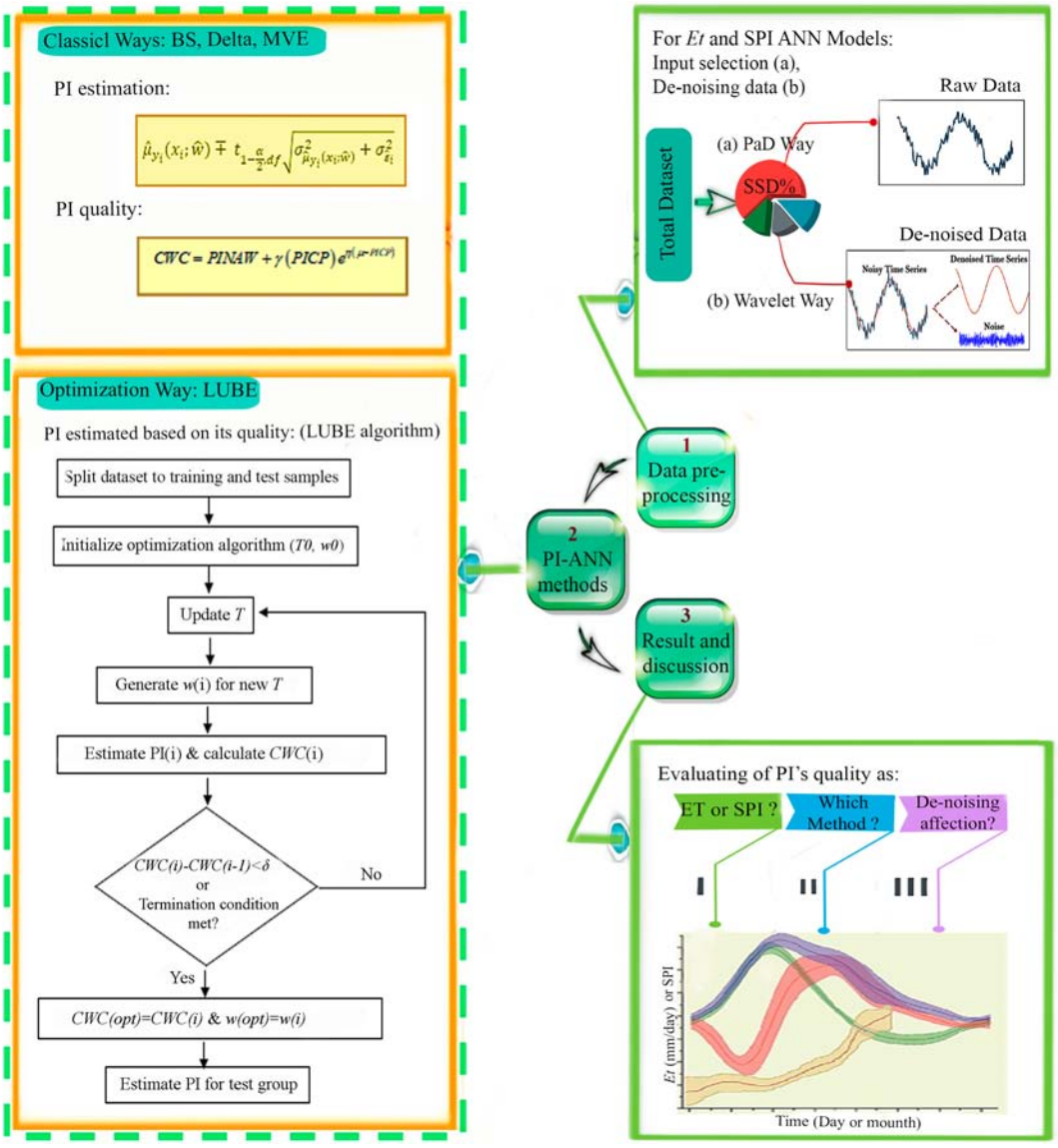

FIG. 2. Proposed methodology scheme. Step 1 involves dominant input selection by the PaD method and then denoising them by the wavelet technique. Step 2 refers to ANN-based PI construction methods (PI-ANN) as classical and optimization (includes LUBE algorithm) methods. Finally, step 3 presents the results of UVs and answers three specific objectives proposed in the introduction section about PI quality. In the LUBE algorithm, T_0 and w_0 are the initial cooling temperature (a parameter of simulated annealing algorithm) and initial ANN parameters; w(opt) is the optimal set of ANN parameters; δ is a number near to zero such as 10^{-6} here.

The current study used PaD and wavelet techniques to eliminate redundant variables and denoise the input data, respectively. Since the ANN model is a kind of regression model, the PaD technique tries to estimate the first-order partial derivatives of the output on the input space (Gevrey et al. 2003; Lu et al. 2001) and the magnitude of the produced PaD values (effectiveness of variables). The sum of squared derivatives (SSD) index is given by Eq. (2) is used to separate dominant (high SSD values) and redundant (low SSD values) variables in the PaD method as follows:

$$\mathrm{SSD}_p = \left[\frac{\sum_{j=1}^{n}\left(\frac{\partial N_o}{\partial X_p}\right)_j^2}{\sum_{i=1}^{\mathrm{ni}}\mathrm{SSD}_i}\right] \times 100\,, \tag{2}$$

where N_o is the output signal from neuron o; X_p is the pth input to the network; and n and ni are the numbers of samples and inputs, respectively. More details about the PaD method can be found in the supplemental material (section S3) as well

as in other reliable sources (e.g., Hashem 1992; Gevrey et al. 2003; Nourani and Sayyah-Fard 2012).

On the other hand, the pattern of ANN output will be more reliable when realistic features of the data are preserved and noise is eliminated. Wavelet-based denoising is suitable for preprocessing and denoising the time series. In this approach, the first step is the decomposition of the time series, i.e., transferring the signal (time series) to the wavelet space. The time series is passed through a series of low-pass and high-pass filters. A detailed signal (or detail coefficients) that includes high-frequency information such as noise is achieved by passing the signal through the high-pass filter. Conversely, the approximation signal (or approximation coefficients) captures low-frequency information by passing the signal through the low-pass filter. This signal includes the identification information of the time series. After the first level of decomposition, detail coefficients are again subjected to wavelet transformation and decomposed into two frequency bounds (high and low coefficients). This process is repeated until the desired frequency resolution is reached. In this way, at the end of the decomposition in the wavelet space, several sets of coefficients are obtained, each set corresponding to a certain frequency range. In the second step, usually the soft thresholding rule (Donoho et al. 1995) according to Eq. (3) is applied to the wavelet coefficients to modify noisy coefficients by filtering some set of them using a given threshold value. Finally, signal reconstruction starts by combining the modified coefficients ($\hat{W}_{j,k}$) in reverse by steps of decomposition from the lowest level upward. Then, denoised time series are used as inputs of the ANN model to estimate the target (here, E_t and SPI). To have the best choice of denoised inputs, which submit high performance of the ANN model, several wavelets with different decomposition levels are tested:

$$\hat{W}_{j,k} = \begin{cases} \text{sgn}(W_{j,k})(|W_{j,k}| - T_j), & |W_{j,k}| > T_j \\ 0, & |W_{j,k}| < T_j \end{cases}, \tag{3}$$

where $W_{j,k}$ and T_j are the detailed coefficient and threshold values, respectively.

To modify the coefficients, universal thresholding (T_j) can be used as follows (Donoho and Johnstone 1994):

$$T_j = \hat{\boldsymbol{\sigma}}\sqrt{2\ln(n)}, \tag{4}$$

$$\hat{\boldsymbol{\sigma}} = \left\{\frac{\text{median}[|W_{j,k}(t)|]}{0.6745}\right\}, \tag{5}$$

where $\hat{\boldsymbol{\sigma}}$ is the noise standard deviation in the noisy time series (Nourani et al. 2014; Sundararajan 2015).

3) Prediction interval estimation by ANN

Uncertainty in the output of a model is a situation where there is no unique and perfectly accurate output. Uncertainty analysis seeks to determine how realistic and reliable the model output of an unknown system is likely to be. ANNs are regression models, and to resolve the uncertainty involved in their results, the regression concept can be applied where the response (or dependent) variable $\mathbf{y}$ is a function of a set of regressors (or independent/explanatory variables) $\{x_1, x_2, \ldots x_n\}$. Targets ($\mathbf{y}$) are related to the inputs ($\mathbf{x}$) by random and deterministic components in the regression analysis where the random components of the target fluctuate around their mean $[\mu_y(\mathbf{x})]$ or deterministic components so that the regression relationship can be written as

$$\mathbf{y} = \mu_y(\mathbf{x}) + \boldsymbol{\varepsilon}, \tag{6}$$

$$\mu_y(\mathbf{x}) = f(\mathbf{x}; w_{\text{true}}), \tag{7}$$

where $\boldsymbol{\varepsilon}$ is the target noise and is assumed to have a Gaussian distribution with zero means, and $\mathbf{w}$ is the set of regression parameters.

The parameters of $\mathbf{w}$ are adjusted as $\hat{w}_0, \hat{w}_1, \ldots \hat{w}_d$ by catching the input data $\{x_1, y_2, \ldots x_n, y_n\}$ pattern to estimate Eq. (7) as

$$\hat{\mu}_y(\mathbf{x}; \hat{\mathbf{w}}) = \hat{w}_o + \sum_{i=1}^{d} \hat{w}_i x_i. \tag{8}$$

In an ANN model, $\hat{\mathbf{w}}$ are weight parameters where due to the random fluctuations of $\mathbf{y}$ around its real mean $[\mu_y(\mathbf{x})]$, the value of $\mathbf{y}$ cannot be obtained with certainty. The best way is to determine the probability density estimation for $\mathbf{y}$, i.e., allocating the area where the contingency of $\mathbf{y}$ is the highest, namely, PIs. When the ANN/regression output $[\hat{\mu}_y(\mathbf{x}; \hat{\mathbf{w}})]$ is subtracted from both sides of Eq. (6), a relationship that illustrates the concepts of UV in mathematical form is obtained as

$$\mathbf{y} - \hat{\mu}_y(\mathbf{x}; \hat{\mathbf{w}}) = [\mu_y(\mathbf{x}) - \hat{\mu}_y(\mathbf{x}; \hat{\mathbf{w}})] + \boldsymbol{\varepsilon}. \tag{9}$$

Confidence intervals (CIs) address the variance of the first term on the right side of Eq. (9) and estimate the uncertainty between the parameters of $\hat{\mu}_y(\mathbf{x}; \hat{\mathbf{w}})$ and $\mu_y(\mathbf{x})$. Instead, PIs seek to quantify the uncertainty between the actual target $\mathbf{y}$ and the estimated amount of $\hat{\mu}_y(\mathbf{x}; \hat{\mathbf{w}})$. This means that targets included random (noise) components that can be effective in the learning process of ANNs (such as other models). With CIs, only the uncertainty of the target's deterministic component is considered during modeling, but using PIs, it is also possible to consider the effect of the uncertainty of the target noise in modeling. In this way, PIs include CIs and are wider.

According to Eq. (9), the whole variance can be written as

$$\sigma^2 = \sigma^2_{\hat{\mu}_y(\mathbf{x};\,\hat{\mathbf{w}})} + \sigma^2_{\hat{\boldsymbol{\epsilon}}}, \tag{10}$$

where $\sigma^2_{\hat{\mu}_y(\mathbf{x};\,\hat{\mathbf{w}})}$ and $\sigma^2_{\hat{\boldsymbol{\epsilon}}}$ are the model and noise variances, respectively (Dybowski and Roberts 2001).

After the total variance of Eq. (10) is obtained, PIs can be estimated via

$$\hat{\mu}_{y_i}(x_i; \hat{\mathbf{w}}) \mp t_{1-(\alpha/2),df}\sqrt{\sigma^2_{\hat{\mu}_{y_i}(x_i;\,\hat{\mathbf{w}})} + \sigma^2_{\hat{\boldsymbol{\epsilon}}_i}}, \tag{11}$$

where $t_{1-(\alpha/2),df}$ is equal to $1 - (\alpha/2)$ cumulative Student's t distribution and freedom degree of df (the number of training samples minus the number of ANN model parameters).

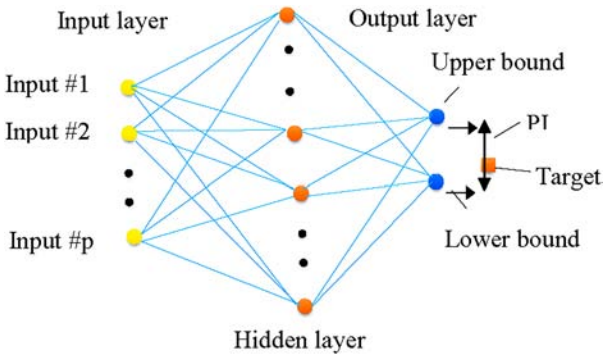

FIG. 3. A typical ANN form used for LUBE implementation to estimate PI_{LUBE}.

Three classical ANN-based methods, which are presented in the supplemental material and section S2, use Eq. (11) to construct PIs by arguing the concepts of the regression model. To estimate PIs using the BS method (PI_{BS}), $\mu_y(\mathbf{x})$ was estimated based on 250 ANNs ($B = 250$), where low values of B can lead to a wide bound, but a greater number of B models also does not affect the quality of PI (Nourani et al. 2022). To estimate PIs using the Delta technique (PI_{Delta}) λ in Eq. (S5) was set to 0.9 (through trial and error), which yielded promising results regarding PI quality.

In addition to classical methods, a new optimization-based method is used to estimate ANN-based PIs, which is described in the following subsection.

4) OPTIMIZATION-BASED LUBE TECHNIQUE

As the optimization-based method, the LUBE is a direct PI construction method based on ANN. For LUBE implementation, unlike common ANNs, there are two outputs as lower and upper bounds of PI_{LUBE} (interval predictions). Such an ANN is trained by a nonlinear cost function named the coverage width criterion (CWC). The CWC combines two conflicting objectives, i.e., PI coverage probability (PICP) and PI normalized average width (PINAW), in which the first should be maximized while the latter should be minimized. PICP is the probability of covering target values with a threshold/nominal confidence value of $(\alpha - 1)\%$. The difference between the upper and lower estimated bounds is measured as PINAW in the normalized form. Because of the complexity of CWC, nongradient optimization methods, i.e., metaheuristic algorithms that do not need the derivatives of the cost function, should be used to solve the proposed optimization problem. Figure 3 illustrates a typical ANN structure used for the LUBE technique. The mathematical formulas of these indices are as follows (Nourani et al. 2019):

$$\mathrm{PICP} = \frac{1}{n}\sum_{i=1}^{n} C_i, \tag{12}$$

$$C_i = \begin{cases} 1, & y_i \epsilon (\hat{y}_{L_i}, \hat{y}_{U_i}) \\ 0, & \text{else} \end{cases}, \tag{13}$$

$$\mathrm{PINAW} = \frac{1}{nR}\sum_{i=1}^{n} (\hat{y}_{U_i} - \hat{y}_{L_i}), \tag{14}$$

TABLE 4. Hyperparameters of the simulated annealing algorithm used in this paper.

Parameters	Numerical value
Geometric cooling schedule	$T_{K+1} = \beta T_K (\beta = 0.9)$
Initial temperature (T0)	5.00
Stopping temperature	0.30
Maximum number of tries within one temperature	30

$$\mathrm{CWC} = \mathrm{PINAW} + \gamma(\mathrm{PICP}) e^{\eta(\mu - \mathrm{PICP})}, \tag{15}$$

$$\gamma(\mathrm{PICP}) = \begin{cases} 1, & \mathrm{PICP} < \mu \\ 0, & \mathrm{PICP} \geq \mu \end{cases}, \tag{16}$$

where $\hat{y}_{L_i}$ and $\hat{y}_{U_i}$ are the upper and lower estimated bounds; y_i is the ith sample of the real target; R is the target interval magnitude, i.e., $y_{max} - y_{min}$; η and μ are two hyperparameters from which the prior is obtained through trial and error and penalizes the CWC for PICP $< \mu$, which causes CWC to grow exponentially and will be controlled, and the latter is equal to $(\alpha - 1)\%$, the nominal confidence value. In this study, η was set equal to 80, in which for its lower values, CWC had more failures to find PICP $\geq \mu$ and μ, as the nominal confidence level was considered 95%.

To construct PI_{LUBE}, the metaheuristic algorithm of simulated annealing was used in this study to minimize the CWC [as well as C_{BS} in Eq. (S3)], in which some of its parameters were selected from related articles and by trial and error, as shown in Table 4. This algorithm is expanded on the basis of the metallurgy annealing process. Annealing is a physical process of heating and slowly cooling the temperature of a material to achieve a uniform crystalline state where the system energy is minimized and thermodynamic equilibrium is established. The simulated annealing optimization algorithm is a free-derivation method and overcomes the disadvantage of the Monte Carlo process in which, without being trapped in local optima, it seeks the parameter space (Kirkpatrick et al. 1983; Aarts and Korst 1989), and its algorithm includes an impressive Metropolis criterion (Metropolis et al. 1953). However, there is a concern when using other metaheuristic algorithms, such as genetic algorithms; it is possible to get stuck in local minima.

At the onset of the optimization process in the LUBE method, PICP $< \mu$; therefore, CWC grows exponentially [see Eq. (15)] without any restrictions for PINAW growth, then the algorithm tries to increase PICP and reduce CWC. After PICP $\geq \mu$ is reached, CWC, and subsequently, PINAW, start to decrease, and ANN parameters are finally optimized and fed into the ANN model with two outputs to build the upper and lower bounds of PI. The lower value of CWC is a sign of the excellent quality of PI, and vice versa; an underestimated PI (PI with lower resolution of 95% confidence) is a result of the high value of CWC.

To evaluate the results obtained by the PI estimation methods, the indicators of PICP, PINAW, and CWC were used according to Eqs. (12), (14), and (15), respectively, and the best results are reported below (in Table 7, for Tabriz station) and illustrated in Figs. 4–7 and S12 (section S7). Based on Fig. 2,

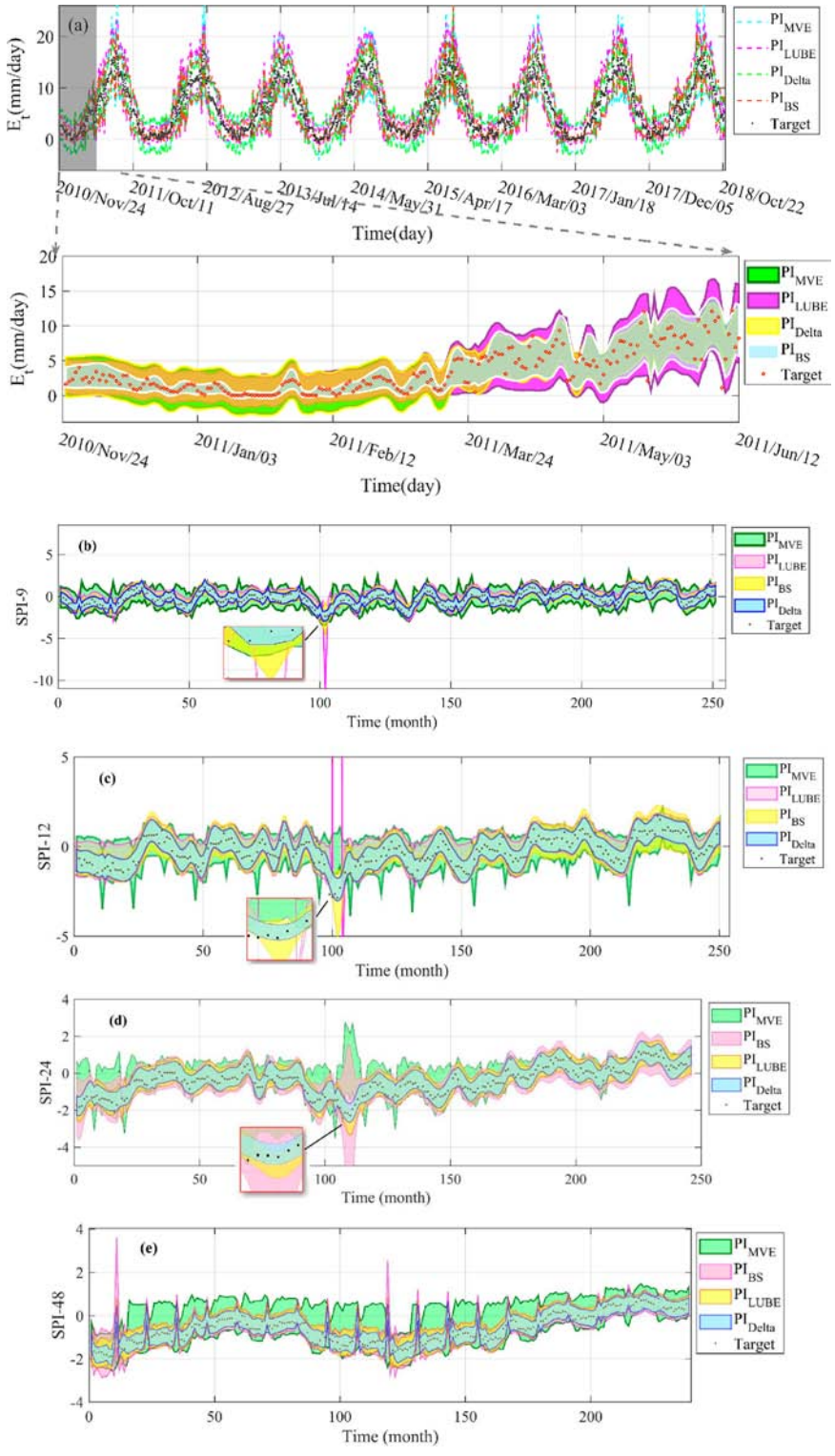

FIG. 4. The estimated PIs using denoised inputs for Tabriz station and models of (a) E_t, (b) SPI-9, (c) SPI-12, (d) SPI-24, and (e) SPI-48.

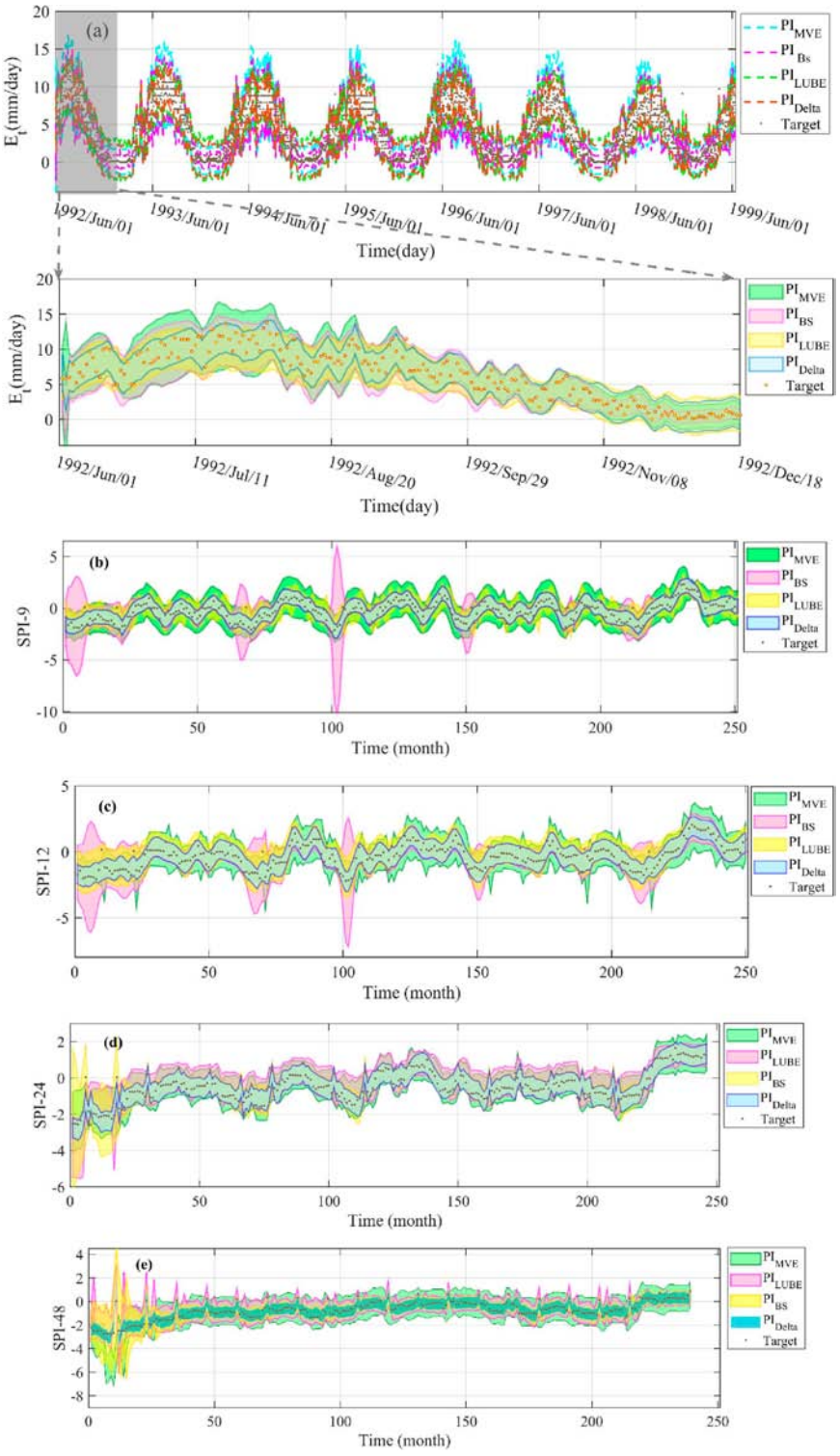

FIG. 5. The estimated PIs using denoised inputs for Urmia station and models of (a) E_t, (b) SPI-9, (c) SPI-12, (d) SPI-24, and (e) SPI-48.

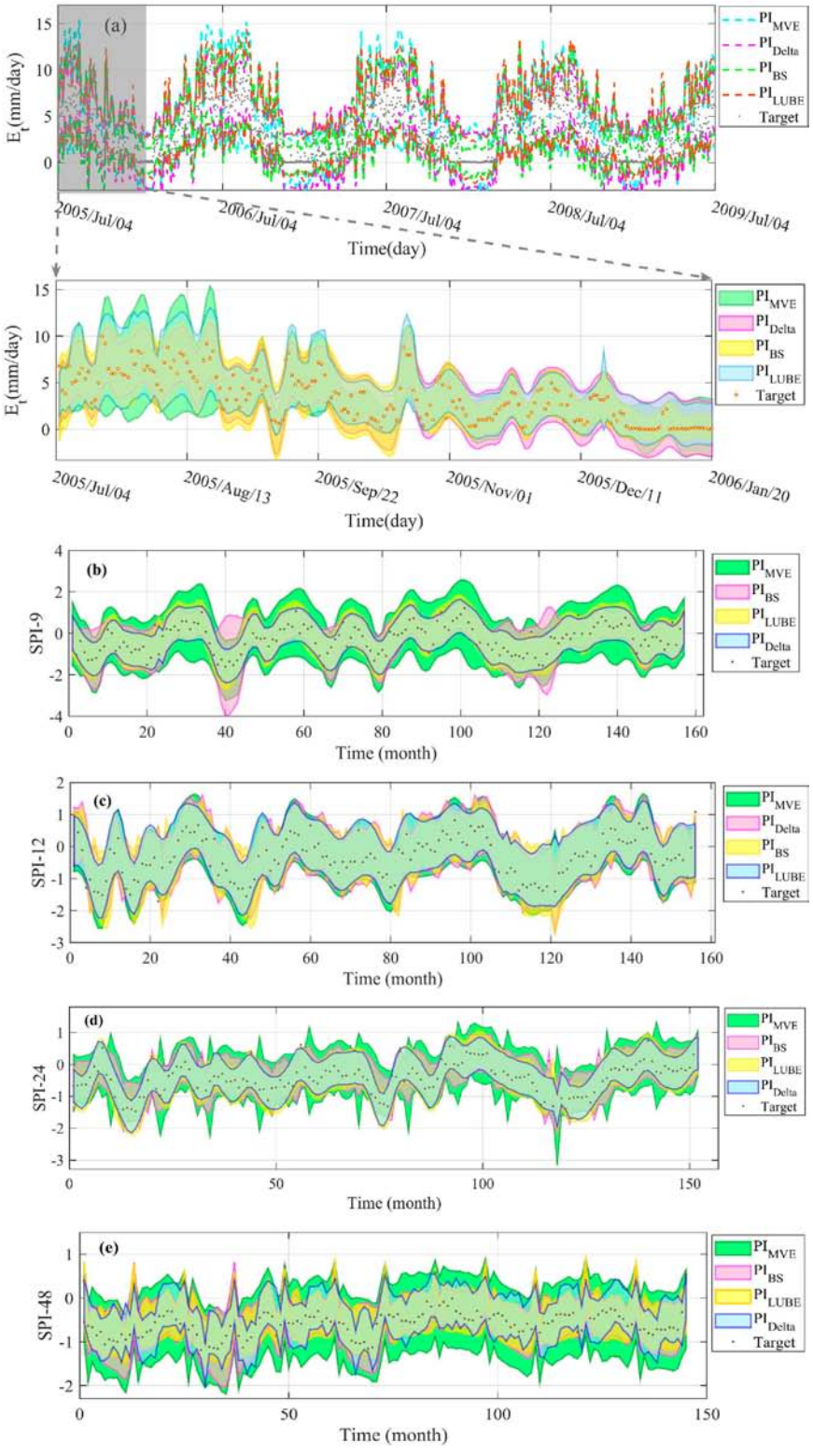

FIG. 6. Estimated PIs using denoised inputs for Ardabil station and models of (a) E_t, (b) SPI-9, (c) SPI-12, (d) SPI-24, and (e) SPI-48.

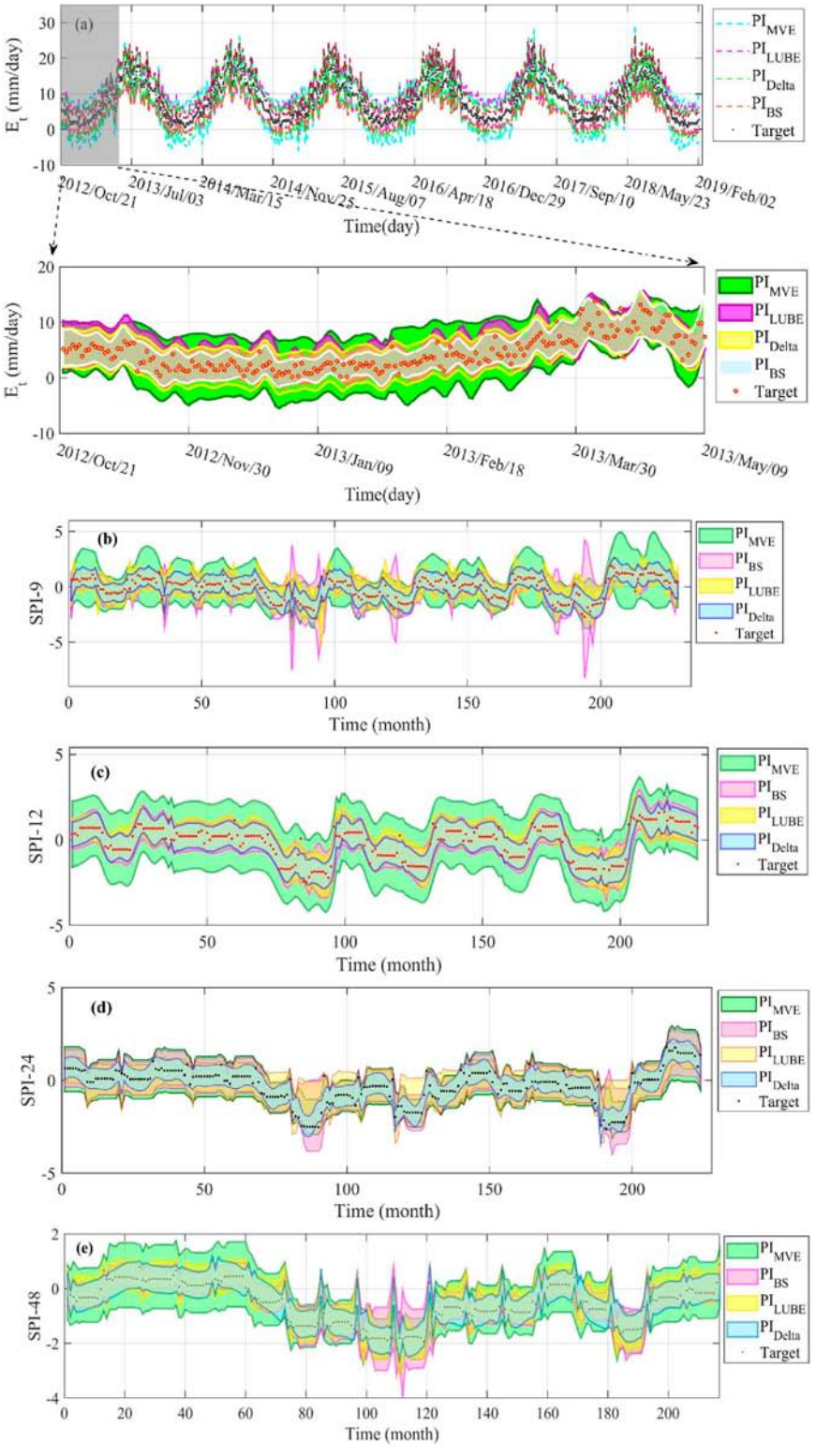

FIG. 7. The estimated PIs using denoised inputs for the Ahvaz station and models of (a) E_t, (b) SPI-9, (c) SPI-12, (d) SPI-24, and (e) SPI-48.

TABLE 5. Dominant input variables in modeling E_t for any station.

Station	Inputs
Tabriz	T, E_{t-1}
Urmia	T, E_{t-1}, R_s, E_{t-2}
Ardabil	T, E_{t-1}, R_h
Ahvaz	T, E_{t-1}, R_h, E_{t-2}

there are three questions about the quality of the calculated PIs, as presented in the following subsections.

3. Results

a. Data preprocessing

At first and since E_t and SPI have the system memory that links them at each time step to the values of previous time steps (lags) as a Markovian (autoregressive) property, potential lags via partial autocorrelation function (PACF) results were detected (see the supplemental material and section S4). Additionally, to enhance the accuracy of SPI modeling, up to four lags of Pr were considered in the total set of potential variables before sensitivity analysis. Then, the dominant inputs were selected by the PaD method among the considered potential inputs with high SSD values (e.g., see Fig. S7), which are tabulated in Tables 5 and 6. Other variables with lower SSD values were removed as redundant variables. It can be seen from Table 5 that for all stations, T and E_{t-1} showed the highest sensitivity to E_t. For Urmia station, and since this city is located in the plain, the selection of R_s as the third most effective input for modeling E_t could be justified. Such topographic areas absorb more R_s because mountains and oceans participate in the formation of clouds as the atmospheric conditions reduce the amount of R_s reaching Earth. Coastal areas influence the Ardabil station due to its proximity to the Astara port, which can be considered a city with a humid climate with a high SSD for the R_h agent.

On the other hand, Ahvaz city can be considered both a windy and humid region. The prevailing winds in this city are the currents that come to Iran from the west, i.e., Iraq and Syria (Broomandi et al. 2017; Abbasi et al. 2021). Additionally, the south wind on the way through the Persian Gulf enters the city of Ahvaz, and in addition to the Karun River, this Gulf is another source of moisture for this city. In addition, it is expected that E_t in Ahvaz city is highly sensitive to R_s values (as Iran's hot climate). Nevertheless, in addition to the low quality of R_s data recorded at the Ahvaz station, dust storms can also be obstacles in reaching the R_s to this city. The sensitivity results for the SPI modeling are also tabulated in Table 6, where regarding the SSD values, up to two or three input variables were dominant for any station. No regulation can be understood that linked the SPI's scale with the selected inputs. The results of this section were used for both point and interval predictions (PIs).

After selecting dominant inputs, the wavelet technique was used to denoise input vectors via the soft thresholding rule and T_j according to Eqs. (3) and (4), and the best BPNN structures for denoised inputs were obtained (presented in section S5 and Table S1). The greatest enhancement was observed in the performance of the SPI models compared to the E_t models. Additionally, among the different time scales, the 9-month and 48-month scales of the SPI showed the most and minor model enhancements, respectively. For the E_t models, the highest increase in performance was related to Ardabil and Urmia stations, and the lowest was related to the Ahvaz station. It seems that the E_t pattern at the Ahvaz station is more regular than that at the other stations, and vice versa, the Urmia and Ardabil stations have the most irregular E_t patterns.

b. Estimation of PIs

1) UNCERTAINTY QUALITY OF PROCESSES

Among all methods, LUBE and BS [using Eq. (11)] yielded acceptable PIs (from CWC, PICP, and PINAW viewpoints), so to interpret the results, PI_{LUBE} and PI_{BS} can be referred to. Despite PI indices, the CV index of observations (CV_O) and UVs (CV_{UV}), obtained from the difference between the upper and the lower bounds or $|U - L|$, can also be used to interpret the results regarding how homogeneous and credible the observations were. For E_t modeling (see Figs. 4a–7a), the narrowest PIs with a CWC equal to 23.93 resulted at Tabriz station, which can be due to the length of the data used as the most extended dataset (26 years). An extensive dataset can help the PI construction algorithms find data patterns with similar confidence (here, μ equals 95%). Furthermore, CV_O and CV_{UV} were close together in modeling E_t at this station, which shows that the UVs of the data were close to each other and that the data were of good quality. After Tabriz station, the low value of CWC was equal to 30.31 at Ahvaz station (with a 21-yr dataset), leading to obtaining narrow PIs in modeling E_t. Additionally, regarding the CV_O and CV_{UV} values as the lowest values of the others, it seems that the pattern of E_t is more homogeneous in this station than in the rest. Conversely, the highest CWC value, equal to 50.14, was obtained for E_t modeling at the Ardabil station, which led to the

TABLE 6. Dominant input variables in modeling SPI for different time scales of the stations.

	Inputs			
Station	SPI-9	SPI-12	SPI-24	SPI-48
Tabriz	SPI_{t-1}, SPI_{t-2}, Pr_{t-1}	SPI_{t-1}, Pr_{t-1}, SPI_{t-2}	SPI_{t-1}, Pr_{t-1}	SPI_{t-1}, Pr_{t-1}, SPI_{t-2}
Urmia	SPI_{t-1}, Pr_{t-1}	SPI_{t-1}, SPI_{t-2}, Pr_{t-1}	SPI_{t-1}, Pr_{t-1}	SPI_{t-1}, SPI_{t-3}, Pr_{t-1}
Ardabil	SPI_{t-1}, Pr_{t-1}	SPI_{t-1}, Pr_{t-1}	SPI_{t-1}, SPI_{t-2}, Pr_{t-1}	SPI_{t-1}, Pr_{t-1}
Ahvaz	SPI_{t-1}, Pr_{t-1}	SPI_{t-1}, Pr_{t-1}	SPI_{t-1}, Pr_{t-1}	SPI_{t-1}, Pr_{t-1}, SPI_{t-2}

TABLE 7. The results of PI estimation methods regarding PI quality at Tabriz station for raw and denoised inputs.

	Raw inputs			Denoised inputs		
Output	CWC	PINAW (%)	PICP (%)	CWC	PINAW (%)	PICP (%)
			BS method			
E_t	25.65	25.65	95.44	23.93	23.93	95.37
SPI-9	43.52	43.52	95.20	34.53	34.53	97.61
SPI-12	45.57	45.57	95.20	37.00	37.00	96.00
SPI-24	103.96	103.96	100.00	48.97	48.97	96.34
SPI-48	63.81	63.81	95.82	38.37	38.37	99.58
			MVE method			
E_t	32.63	32.63	95. 95	30.40	30.40	97.00
SPI-9	74.40	74.40	98.40	55.67	55.67	98.80
SPI-12	55.28	55.28	98.79	45.82	45.82	98.00
SPI-24	54.87	54.87	99.59	43.07	43.07	99.18
SPI-48	62.13	62.13	98.74	55.64	55.64	98.32
			Delta method			
E_t	31.23	31.23	95.20	27.08	27.08	96.20
SPI-9	40.83	40.83	95.22	26.66	26.66	96.81
SPI-12	568.47	35.13	91.60	181.97	26.71	92.80
SPI-24	597.30	33.31	92.68	202.78	27.47	92.68
SPI-48	123.31	50.93	94.56	24.97	24.97	95.82
			LUBE method			
E_t	30.90	30.90	96.95	27.74	27.74	95.92
SPI-9	58.33	58.33	95.62	36.31	36.31	96.41
SPI-12	44.89	44.89	95.60	37.44	37.44	96.80
SPI-24	45.76	45.76	96.34	35.88	35.88	95.93
SPI-48	51.18	51.18	96.65	33.82	33.82	97.49

widest PIs. A significant discrepancy between CV_O and CV_{UV} values is observed at this station, equal to 0.45%, which seems that the PIs could not be very well associated with the UV of the data at this station. At the Urmia station, the best CWC value was 32.85, which showed a good state of PI regarding UV, but the most anomalous E_t and the associated UV patterns were obtained at this station.

Figures 4–7 also illustrate the PIs for the SPI models, while Table 7 presents the UV results for the Tabriz station. At this station, some extreme drought events of ED (SPI ≤ -2) for the SPI-9 and SPI-12 models (which are outstanding in Figs. 4b,c) were not covered by PI_{LUBE} but were covered by other PIs. The PI_{LUBE} for these SPI values jumped with PINAW near zero, but the PI_{BS} covered them by widening the PINAW. Due to the dissimilar behavior of different PIs for these extreme events, they cannot be concluded in terms of 95% confidence. From the PI_{BS} and PI_{LUBE} perspectives (see Figs. 4b–e), at Tabriz station, a low CWC value of 34.53 was obtained for the SPI-9 model. Additionally, the lowest CV_{UV} was received for this time scale, equal to 0.12%, therefore, it can be interpreted that the UVs for the SPI-9 data were close to each other and for monitoring the meteorological drought at this station, the 9-month scale is more certain and recommended. For models of SPI-24 and SPI-48 at the Uemia station, some extreme events of ED and SD ($-1.99 \leq$ SPI ≤ -1.50) obtained high UV in the $PI_{95\%}$ and may be sticking out from the PIs by reducing the amount of μ in the modeling of SPI-24 and SPI-48. However, the lowest CWC for the SPI-24 model, equal to 32.21 at this station, resulted in narrow PIs (see Figs. 5b–e). Additionally, the lowest CV_{UV} was obtained for the output of the SPI-24 models, equal to 0.19%, where CV_O was also low for this time scale, indicating low dispersion of the SPI-24 data. It could be interpreted that the UVs of the data on the 24-month scale were close to each other, and they were more homogeneous and can be recommended for meteorological drought monitoring at Urmia station.

Among all stations, the lowest number of extreme events emerged in the DIP output for Ardabil station. Additionally, the short length of data in modeling SPIs at Ardabil station can affect the quality of the estimated PIs (see Figs. 6b–e) via the gamma distribution perspective. This issue has already been maintained by Wang et al. (2021). This means that estimating gamma distribution parameters (shape and scale parameters) is sensitive to data size. For a minimal data length, they can be more unstable/uncertain. The lowest CWC at this station was 48.42 for model SPI-24, and the lowest CV_{UV} was also obtained for this model. From the PI indices, CV_O, and CV_{UV} perspectives, data at the 24-month scale at Ardabil station are more certain and homogenous; therefore, this scale can be recommended to monitor meteorological drought. For the Ahvaz station, the results of the PIs for the SPI models are presented in Figs. 7b–e. From the results, a low CV_{UV} equal to 0.13% was obtained for the SPI-24 model, while CV_O was also low for this scale, which can be interpreted as homogeneity of the data. Furthermore, a low CWC value of 35.91 for this model represented the narrow PIs; therefore, this scale can be recommended for monitoring meteorological drought at the Ahvaz station.

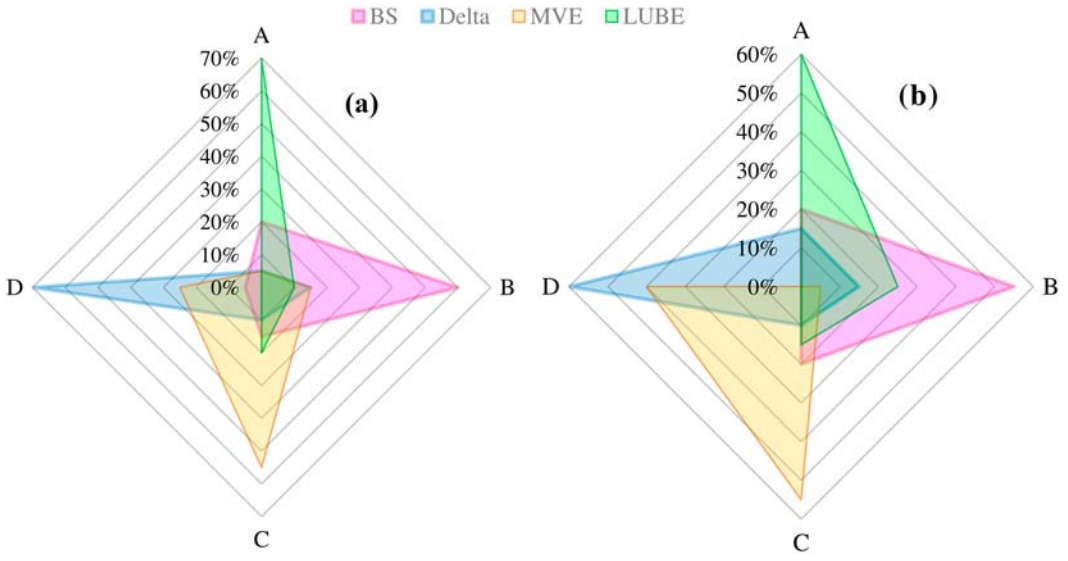

FIG. 8. Radar plots of the contribution of four methods to evaluate UV in all models. Ranks of A, B, C, and D were achieved based on the CWC values for (a) raw input data and (b) denoised input data.

2) COMPARISON OF DIFFERENT PI CONSTRUCTION METHODS

From the results, the greatest similarity in terms of PI swings was established for PI_{BS}, PI_{Delta}, and PI_{LUBE} for all stations, but PI_{MVE} showed the opposite swings compared to the rest in some cases (see Figs. S8–S11, section S6). In most PIs, the PINAW values for PI_{Delta} [using Eq. (S5)] and PI_{MVE} [using Eq. (11)] were the lowest and highest, respectively (see Figs. 4–7 and Table 7). From the mathematical viewpoint, PINAW is more influenced by σ_ε^2; therefore, this parameter was underestimated and overestimated by PI_{Delta} and PI_{MVE}, respectively (as maintained already by Khosravi et al. 2011a,b). Additionally, it is clear from the results that PI_{BS} and PI_{LUBE} have better conveyed the uncertainty of the modeling. Radar charts in Fig. 8 show the rank of four methods in evaluating UV in all models, based on CWC, when raw and denoised input data were used. From the results, the PI_{LUBE}, in most cases, ranked first in terms of the lowest CWC values (corresponding to the lowest PINAWs for PICP $\geq \mu$); therefore, the LUBE method can be considered the most robust method to estimate PIs, which was also pointed out in previous studies (e.g., Khosravi et al. 2011a,b; Nourani et al. 2019, 2022). Unlike the PIs_{LUBE}, PIs_{Delta} obtained the highest CWC values in most cases and ranked fourth. Delta technique failed to find PICP $\geq \mu$ for 26 cases out of the total raw and denoised data due to its narrow PINAWs (underestimated PIs). From the radar charts, after the LUBE method, BS was the robust technique for determining the UV of the outputs and ranked second.

3) THE EFFECT OF DENOISING ON THE QUALITY OF PIS

In all cases for E_t and SPI modeling, imposing the denoised input data led to a reduction in the UV of the output, which has been manifested in reducing CWCs and PINAWs. The highest reductions of PIs_{BS} and PIs_{LUBE} are tabulated in Table 8. The amount of uncertainty reduction depends on the technique used and the input vector size, data quality, and station climate (in the form of dominant inputs selected for modeling). The lowest reduction in the output UV was related to E_t modeling, which indicates the lower randomness of this process compared to drought (SPI). From the perspectives of PI_{BS} and PI_{LUBE} and according to Table 8, the highest reduction of output UV was related to SPI-9, equal to 76%, and PI_{BS} at the Urmia station. Additionally, for E_t modeling, the highest reduction was 30% at the Ardabil station when estimating the PIs_{BS}. Figure S12 (section S7) illustrates the PIs for the highest reduction of the SPI and E_t models.

TABLE 8. The highest reductions in output UV from the PI_{BS} and PI_{LUBE} perspectives using denoised data at any station.

	Δ(CWC) (%)		Δ(CWC) (%)	
Output	PI_{LUBE}	PI_{BS}	PI_{LUBE}	PI_{BS}
	Tabriz		Urmia	
E_t	10.23	6.71	17.17	18.08
SPI-9	37.75	20.66	58.51	76.01
SPI-12	16.60	18.81	16.60	20.07
SPI-24	21.59	52.90	7.58	24.50
SPI-48	33.92	39.87	56.21	50.39
	Ardabil		Ahvaz	
E_t	15.98	30.06	16.62	0.82
SPI-9	18.60	7.88	10.62	8.88
SPI-12	11.06	5.86	16.61	14.51
SPI-24	18.32	16.05	17.05	27.56
SPI-48	9.68	18.11	9.48	30.54

4. Discussion

a. *Effect of high-frequency components and data length on the quality of PIs*

Concerning the results, the type of phenomena, the data quality, and the stations' climatic regime in the form of selected inputs affect the quality of the estimated PIs. For a target variable (E_t or SPI), the change in station led to a change in the ranking of all methods regarding the quality of the obtained PIs. Additionally, the ranking of the PI construction methods was altered for a desired station but different target variables (E_t or SPI). Such a discrepancy can be interpreted because uncertainty analysis is rooted in the random nature of events. Namely, it is clear that the randomness degree of Pr (the magnitude of high-frequency components in the Pr time series) is more than that of E_t and therefore is subjected to more uncertainty where SPI models obtained lower quality PIs (in all of its scales) than E_t. Another interpretation for the obtained PIs in terms of the quality can be related to "the time scale effection." This means that on the daily time scale, due to the large data length, identifying the pattern of data for the ANN model is more accurate than on the monthly scale. The Ardabil station participated in the whole modeling with the lowest data length, which achieved poor PIs. It seems that increasing data length allows the PI estimation algorithm to distinguish the available pattern of the data better.

b. *Different methods submitted different resolutions in estimating PIs*

Different methods submitted different resolutions in estimating PIs. In the category of classical methods, it seems that the BS method has performed better than others. From the BS perspective, it seems that B ANN models, trained with various random weight vectors, could better cover the space of the network parameters. Aggregating the results in the BS method could lead to an accurate estimation of the target average. Compared to the $\mathrm{PI}_{\mathrm{LUBE}}$ and $\mathrm{PI}_{\mathrm{Delta}}$, the $\mathrm{PI}_{\mathrm{BS}}$ was wider; this expanded width can be related to the overestimated total variance. In the present study, to reduce PIVAWs, the number of B was set to 250 through trial and error. There is an inverse relationship (negative) between the number of B networks and the CWC value, but this relation is not strong, and a large B does not guarantee good quality of PIs (by reducing CWC). $\mathrm{PIs}_{\mathrm{MVE}}$ in most cases ranked third in terms of the CWC index. Although the $\mathrm{PIs}_{\mathrm{MVE}}$ were better than the $\mathrm{PIs}_{\mathrm{Delta}}$ and included high PICPs, due to their wide PINAWs (overestimated PIs), they could not convey helpful information about uncertainty. One source for the low quality of $\mathrm{PIs}_{\mathrm{MVE}}$ could be linked to its basic assumption: its model variance is negligible as zero, but this cannot be correct in real-world practical applications. From Figs. S8–S11, the $\mathrm{PI}_{\mathrm{Delta}}$ shows the lowest fluctuations and, in most cases, especially for models of SPI, did not accompany the UV of the observations and resulted in PICP $< \mu$. However, the Delta method showed almost reliable results for E_t modeling by providing PICP $\geq \mu$. Additionally, the Delta technique in the Urmia station and confronting E_t and SPI mostly showed weak performance. It should be noted that the data of the Urmia station include irregular patterns of E_t. However, $\mathrm{PIs}_{\mathrm{BS}}$ and $\mathrm{PIs}_{\mathrm{LUBE}}$, as two methods with acceptable results, had swung well with data points, which denotes a suitable convey of UV of the observations by them. The LUBE method, which does not include any assumptions to estimate PIs and is an optimization algorithm, can be recommended as a robust technique for PI estimation.

c. *Drought risk from PI's (uncertainty science) perspective*

As a whole, it can be interesting to have an initial guess about drought risk from the uncertainty science perspective knowing that meteorological droughts can lead to hydrological and agricultural droughts with a lagged time. This means that a lack of soil moisture or loss of surface water can confront terrestrial ecosystems with water stress and propagate through the hydrological cycle. The results obtained for the Tabriz, Urmia, Ardabil, and Ahvaz station SPI models at 9, 24, 24, and 24 months showed low UV and narrow PIs. These time scales were recommended to monitor meteorological drought at relevant stations. As the time scale becomes larger, the propagation risk of meteorological drought to hydrological or agricultural drought (hydrological cycle) increases. From this viewpoint, Tabriz station, with the recommended time scale to monitor meteorological drought (9 months), is the lowest-risk station.

Despite the hot climate, Ahvaz city is also considered humid. In this city, the rate of E_t in the hot period (spring and summer) was less than that at the Tabriz and Urmia stations and almost equal to that at the Ardabil station. This issue was unexpected due to the hot climate of the Ahvaz station. Nevertheless, according to previous studies that proved the existence of the E_t paradox (e.g., Nasrollahi et al. 2021), it can be concluded that E_t includes a complex but homogeneous pattern at this station. Changes in energy equilibrium make the rise in temperature expected to prepare more energy to evaporate water sources, accelerating the hydrological cycle. Nevertheless, paradoxically, the E_t gradient reverses, which is named the E_t paradox. The E_t's reverse gradient can occur due to various climatic agents. It seems that one of the driving climatic factors is the humidity of the station in the hot period. The inverse rate of E_t can help to increase the soil moisture (or, in its high values, can cause scattered rainfall and increase surface water), so it can be said that this phenomenon can reduce the propagation risk of meteorological drought through the hydrological cycle even for a 24-month scale, which was recommended in this study for drought monitoring. From this perspective, no strong correlation between meteorological drought and other types of droughts at this station is conceivable. It should be emphasized that because this city has a hot/humid climate, monitoring the drought at this station using the multivariate index, which also applies the E_t (the SPI only considers Pr), such as the SPEI, can be more informative than the SPI. Even though the multivariable indices take

into account more details, the anomalies of Pr and E_t, such as when using SPEI, can affect each other; therefore, there is no guarantee to reduce the uncertainty in using such an index.

At the Ardabil station, despite the more certain model of SPI-24 to monitor meteorological drought, the lowest E_t values (in the warm periods) and humidity agents at the Ardabil station can decrease the propagation risk of meteorological drought. However, at the Urmia station, confidence in the SPI-24 model and a high rate of E_t in the hot periods can significantly correlate meteorological drought with hydrological/agricultural drought. From this perspective, the drought problem can be severe for Urmia station. It should be noted that meteorological drought can link to other types of drought, but its monitoring indices cannot be used to monitor different kinds of drought.

5. Conclusions

In the present study, PIs were estimated for ANN-based modeling of E_t and SPI using conventional techniques of MVE, BS, Delta, and an optimization-based method named LUBE. From the results, the LUBE and BS techniques were two suitable tools for estimating PIs, but the LUBE method was superior in its assumption-free and robust solution. In the majority of cases, the MVE and Delta techniques led to an overestimation and underestimation of PIs. We investigated the effect of high-frequency components of the target (E_t or SPI) time series (higher disturbance or the randomness of target) on the PIs. From the results, PIs for the E_t models included the lowest CWC values up to 50.14 (at the Ardabil station), while the SPI models had wider PIs with CWC values up to 64.75 (at the Ardabil station). It means the abnormal high-frequency components of the SPI time series led to wider PIs than those for the E_t. Additionally, it seems that the short data length led to poor PIs, which was clear for the estimated PIs with the lowest data length at the Ardabil station. To enhance the quality of the PIs, a hybrid methodology including PaD and wavelet techniques was linked to the uncertainty analysis to remove respectively the redundant inputs and noise components (high-frequency components of the input vector) which could lead to narrower PIs with more significant effects on the SPI models (up to 76%).

From a high-quality PI perspective, SPI-9, SPI-24, SPI-24, and SPI-24 for the Tabriz, Urmia, Ardabil, and Ahvaz stations were recommended in this study for monitoring meteorological drought. Additionally, from weather and PI (uncertainty) viewpoints, it is possible to make a primary sense for the drought risk. It seems that the risk of propagation of meteorological drought through the hydrological cycle at Urmia station is more severe than that at the other stations. The effect of other distribution type functions on the quality of PIs in estimating SPIs (rather than gamma distribution) can be investigated in future studies. The current research dedicated a separate viewpoint to SPI and E_t to investigate the differences between these two processes regarding their uncertainty in different climatic regions. However, research can be continued by catching them in one framework as a multivariate index such as SPEI. To achieve optimal PIs, linear and nonlinear ensemble methods that combine different PIs can be investigated.

Acknowledgments. This work was supported by the core-to-core project, the Japan Society for the Promotion of Science (JSPS) (Grant JPJSCCB202200044), and the Disaster Prevention Research Institute (DPRI) internal research fund (Grant 2022 L-04).

Data availability statement. The data supporting the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX A

List of Abbreviations

Items with an asterisk are in the supplemental material.

T	Temperature
E_t	Evaporation
E_{t-1}	One-day lagged evaporation
R_s	Solar radiation
P	Air pressure
R_h	Relative humidity
U	Wind speed
Pr	Precipitation
SPI	Standardized precipitation index
SPEI	Standardized precipitation evapotranspiration index
ANN	Artificial neural networks
BPNN	Backpropagation ANN
BS	Bootstrap
MVE	Mean variance estimation
LUBE	Lower upper bound estimation
UV	Uncertainty value
PI	Prediction interval
CI	Confidence interval
DIP	Drought indices package
PaD	Partial derivatives
RMSE	Root-mean-square error
MSE	Mean squared error
DC	Determination coefficient
SSD	Sum of squared derivatives
PI_{LUBE}	PI-based LUBE
PI_{BS}	PI-based Bs
PI_{MVE}	PI-based MVE
PI_{Delta}	PI-based Delta
CWC	Coverage width criterion
C_{BS}*	The cost function for optimizing ANNσ parameters
PICP	PI coverage probability
PIVAW	PI normalized average width
PACF	Partial autocorrelation function
CV_{UV}	CV index of uncertainty values
CVo	CV index of observations
$\|U - L\|$	Difference between the upper and lower bounds
symN*	Symlet wavelet function symbol
dbN*	Daubechies wavelet function symbol

ANN_{σ}*	An ANN model to estimate target noise variances
ANN_{y}*	An ANN model to estimate the mean of the target

APPENDIX B

List of Symbols

Items with an asterisk are in the supplemental material.

N_o	The output signal from neuron o
X_p	The pth input to the network
n	Number of samples
ni	Number of inputs
$W_{j,k}$	The detailed coefficient
$W_{\hat{j},k}$	The modified coefficient
T_j	Universal threshold value
$\hat{\sigma}$	The noise standard deviation [Eq. (4)] of time series (here, input vector)
$\mathbf{y}$	Target vector
$\mathbf{x}$	Input vector
$\mu_y(\mathbf{x})$	Mean of the target
$\boldsymbol{\varepsilon}$	Target noise
$\mathbf{w}$	The set of regression parameters (or ANN weight vector)
$\hat{\mathbf{w}}$	Estimated weight vector
$\hat{\mu}_y(\mathbf{x}; \hat{\mathbf{w}})$	Estimated mean of the target or ANN/regression output
σ^2	Total ANN output variance [as described in Eq. (9)]
$\sigma^2_{\hat{\mu}_y(\mathbf{x}; \hat{\mathbf{w}})}$	Model variance
$\sigma^2_{\hat{\boldsymbol{\varepsilon}}}$	Noise variances of the target
$t_{1-(\alpha/2),df}$	Equal to $1 - (\alpha/2)$ cumulative Student's t distribution and freedom degree of df
$\hat{y}_{L_i}$	The upper estimated bound
$\hat{y}_{U_i}$	The lower estimated bound
y_i	The ith sample of the real target
R	Target interval magnitude ($y_{\max} - y_{\min}$)
η	A hyperparameter to control CWC
μ	The nominal confidence value and is equal to $(\alpha - 1)\%$
λ	Regularization parameter
d*	The number of ANN parameters
r_i^2*	A set of residuals of variance squared in BS and MVE methods
$\hat{\mu}^b_{y_i}(x_i; \hat{\mathbf{w}})$*	The ith point sample predicted by the bth model in the BS method
w^**	The set of optimal ANN parameters
g_0^T*	The output gradient for parameters of w^* and new dataset
$\mathbf{F}$*	The Jacobian matrix
O*	The new dataset (test data)

REFERENCES

Aarts, E., and J. Korst, 1989: *Simulated Annealing and Boltzmann Machine: A Stochastic Approach to Combinatorial Optimization and Neural Computing.* John Wiley and Sons, 272 pp.

Abbasi, H. R., C. Opp, M. Groll, A. Gohardoust, and H. Rouhipour, 2021: Wind regime and aeolian sand transport in Khuzestan Sand Sea. *Aeolian Res.*, **53**, 100746, https://doi.org/10.1016/j.aeolia.2021.100746.

Aryal, A., M. Maharjan, R. Talchabhadel, and B. R. Thapa, 2022: Characterizing meteorological droughts in Nepal: A comparative analysis of Standardized Precipitation Index and rainfall anomaly index. *Earth*, **3**, 409–432, https://doi.org/10.3390/earth3010025.

Azzahari, A. D., S. N. F. Yusuf, V. Selvanathan, and R. Yahya, 2016: Artificial neural network and response surface methodology modeling in ionic conductivity predictions of Phthaloylchitosan-based gel polymer electrolyte. *Polymers*, **8**, 22, https://doi.org/10.3390/polym8020022.

Broomandi, P., B. Dabir, B. Bonakdarpour, and Y. Rashidi, 2017: Identification of the sources of dust storms in the City of Ahvaz by HYSPLIT. *Pollution*, **3**, 341–348.

Campisi-Pinto, S., J. Adamowski, and J. Oron, 2012: Forecasting urban water demand via wavelet-denoising and neural network models. Case study: City of Syracuse, Italy. *Water Resour. Manage.*, **26**, 3539–3558, https://doi.org/10.1007/s11269-012-0089-y.

Chen, Y., L. Song, Y. Liu, L. Yang, and D. Li, 2020: A review of the artificial neural network models for water quality prediction. *Appl. Sci.*, **10**, 5776, https://doi.org/10.3390/app10175776.

de Veaux, R. D., J. Schumi, J. Schweinsberg, and L. H. Ungar, 1998: Prediction intervals for neural networks via nonlinear regression. *Technometrics*, **40**, 273–282, https://doi.org/10.2307/1270528.

Donoho, D. L., and I. M. Johnstone, 1994: Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**, 425–455, https://doi.org/10.1093/biomet/81.3.425.

——, ——, G. Kerkyacharian, and D. Picard, 1995: Wavelet shrinkage: Asymptopia? *J. Roy. Stat. Soc.*, **57B**, 301–337, https://doi.org/10.1111/j.2517-6161.1995.tb02032.x.

Dybowski, R., and S. Roberts, 2001: Confidence intervals and prediction intervals for feed-forward neural networks. *Clinical Applications of Artificial Neural Networks*, Cambridge University Press, 298–326, https://doi.org/10.1017/CBO9780511543494.013.

Efron, B., 1979: Bootstrap methods: Another look at the jackknife. *Ann. Stat.*, **7** (1), 1–26, https://doi.org/10.1214/aos/1176344552.

Gedeon, T. D., 1997: Data mining of inputs: Analysing magnitude and functional measures. *Int. J. Neural Syst.*, **8**, 209–218, https://doi.org/10.1142/S0129065797000227.

Gevrey, M., I. Dimopoulos, and S. Lek, 2003: Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecol. Modell.*, **160**, 249–264, https://doi.org/10.1016/S0304-3800(02)00257-0.

——, ——, and ——, 2006: Two-way interaction of input variables in the sensitivity analysis of neural network models. *Ecol. Modell.*, **195**, 43–50, https://doi.org/10.1016/j.ecolmodel.2005.11.008.

Guo, J., J. Zhou, H. Qin, Q. Zou, and Q. Li, 2011: Monthly stream-flow forecasting based on improved support vector machine model. *Expert Syst. Appl.*, **38**, 13 073–13 081, https://doi.org/10.1016/j.eswa.2011.04.114.

Hashem, S., 1992: Sensitivity analysis for feed forward artificial neural networks with differentiable activity functions. *Int. Conf. on Neural Network*, Baltimore, MD, Institute of Electrical and Electronics Engineers, 419–429, https://doi.org/10.1109/IJCNN.1992.287175.

Hwang, J. T. G., and A. A. Ding, 1997: Prediction intervals for artificial neural networks. *J. Amer. Stat. Assoc.*, **92**, 748–757, https://doi.org/10.1080/01621459.1997.10474027.

Kabir, D. H. M., A. Khosravi, M. Anwar Hosen, and S. Nahavandi, 2018: Neural network-based uncertainty quantification: A survey of methodologies and applications. *IEEE Access*, **6**, 36 218–36 234, https://doi.org/10.1109/ACCESS.2018.2836917.

Kamruzzaman, M., and Coauthors, 2022: Spatiotemporal drought analysis in Bangladesh using the Standardized Precipitation Index (SPI) and standardized precipitation evapotranspiration index (SPEI). *Sci. Rep.*, **12**, 20694, https://doi.org/10.1038/s41598-022-24146-0.

Kasiviswanathan, K. S., R. Cibin, K. Sudheer, and I. Chaubey, 2013: Constructing prediction interval for artificial neural network rainfall runoff models based on ensemble simulations. *J. Hydrol.*, **499**, 275–288, https://doi.org/10.1016/j.jhydrol.2013.06.043.

Khosravi, A., S. Nahavandi, D. Creighton, and A. F. Atiya, 2011a: Lower upper bound estimation method for construction of neural network-based prediction intervals. *IEEE Trans. Neural Networks*, **22**, 337–346, https://doi.org/10.1109/TNN.2010.2096824.

——, ——, ——, and ——, 2011b: Comprehensive review of neural network-based prediction intervals and new advances. *IEEE Trans. Neural Networks*, **22**, 1341–1356, https://doi.org/10.1109/TNN.2011.2162110.

Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi, 1983: Optimization by simulated annealing. *Science*, **220**, 671–680, https://doi.org/10.1126/science.220.4598.671.

Komuscu, A. U., 1999: Using the SPI to analyze spatial and temporal patterns of drought in Turkey. *Drought Network News*, Vol. 11, No. 1, University of Nebraska–Lincoln, Lincoln, NE, 8 pp., https://digitalcommons.unl.edu/droughtnetnews/49/

Koutroulis, A. G., A. K. Vrohidou, and I. K. Tsanis, 2011: Spatiotemporal characteristics of meteorological drought for the island of Crete. *J. Hydrometeor.*, **12**, 206–226, https://doi.org/10.1175/2010JHM1252.1.

Kumar, M. N., C. S. Murthy, M. V. R. Sesha Sai, and P. S. Roy, 2009: On the use of Standardized Precipitation Index (SPI) for drought intensity assessment. *Meteor. Appl.*, **16**, 381–389, https://doi.org/10.1002/met.136.

Kumar, S., M. K. Tiwari, C. Chatterjee, and A. Mishra, 2015: Reservoir inflow forecasting using ensemble models based on neural networks, wavelet analysis and bootstrap method. *Water Resour. Manage.*, **29**, 4863–4883, https://doi.org/10.1007/s11269-015-1095-7.

Li, Z., and Coauthors, 2021: Clarifying the propagation dynamics from meteorological to hydrological drought induced by climate change and direct human activities. *J. Hydrometeor.*, **22**, 2359–2378, https://doi.org/10.1175/JHM-D-21-0033.1.

Lian, C., Z. Zeng, W. Yao, H. Tang, and C. L. P. Chen, 2016: Landslide displacement prediction with uncertainty based on neural networks with random hidden weights. *IEEE Trans. Neural Networks Learn. Syst.*, **27**, 2683–2695, https://doi.org/10.1109/TNNLS.2015.2512283.

Lu, M., S. M. Abourizk, and U. H. Hermann, 2001: Sensitivity analysis of neural networks in spool fabrication productivity studies. *J. Comput. Civ. Eng.*, **15**, 299–308, https://doi.org/10.1061/(ASCE)0887-3801(2001)15:4(299).

McKee, T. B., N. J. Doesken, and J. Kleist, 1993: The relationship of drought frequency and duration to time scales. *Eighth Conf. on Applied Climatology*, Anaheim, CA, Amer. Meteor. Soc., 179–184.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, 1953: Equation of state calculation by fast computing machines. *J. Chem. Phys.*, **21**, 1087–1092, https://doi.org/10.1063/1.1699114.

Nasrollahi, M., A. A. Zolfaghari, and M. R. Yazdani, 2021: Investigation of pan evaporation paradox and climatic parameters affecting it in half-west and center of Iran (in Persian). *J. Water Soil Resour. Conserv.*, **11**, 61–76, https://doi.org/10.30495/wsrcj.2021.18545.

Nix, D. A., and A. S. Weigend, 1994: Estimating the mean and variance of the target probability distribution. *Proc. 1994 IEEE Int. Conf. on Neural Networks (ICNN'94)*, Orlando, FL, Institute of Electrical and Electronics Engineers, 55–60, https://doi.org/10.1109/ICNN.1994.374138.

Nourani, V., and M. Sayyah-Fard, 2012: Sensitivity analysis of the artificial neural network outputs in simulation of the evaporation process at different climatologic regimes. *Adv. Eng. Software*, **47**, 127–146, https://doi.org/10.1016/j.advengsoft.2011.12.014.

——, A. H. Baghanam, J. Adamowski, and O. Kisi, 2014: Applications of hybrid wavelet–artificial intelligence models in hydrology: A review. *J. Hydrol.*, **514**, 358–377, https://doi.org/10.1016/j.jhydrol.2014.03.057.

——, N. J. Paknezhad, E. Sharghi, and A. Khosravi, 2019: Estimation of prediction interval in ANN-based multi-GCMs downscaling of hydro-climatologic parameters. *J. Hydrol.*, **579**, 124226, https://doi.org/10.1016/j.jhydrol.2019.124226.

——, M. Sayyah-Fard, M. T. Alami, and E. Shargi, 2020: Data pre-processing effect on ANN-based prediction intervals construction of the evaporation process at different climate regions in Iran. *J. Hydrol.*, **588**, 125078, https://doi.org/10.1016/j.jhydrol.2020.125078.

——, H. Najafi, E. Sharghi, and K. Roushangar, 2021: Application of Z-numbers to monitor drought using large-scale oceanic-atmospheric parameters. *J. Hydrol.*, **598**, 126198, https://doi.org/10.1016/j.jhydrol.2021.126198.

——, K. Khodkar, and M. Gebremichaelc, 2022: Uncertainty assessment of LSTM based groundwater level predictions. *Hydrol. Sci. J.*, **67**, 773–790, https://doi.org/10.1080/02626667.2022.2046755.

Pande, C. B., N. Al-Ansari, N. L. Kushwaha, A. Srivastava, R. Noor, M. Kumar, K. N. Moharir, and A. Elbeltagi, 2022: Forecasting of SPI and meteorological drought based on the artificial neural network and M5P model tree. *Land*, **11**, 2040, https://doi.org/10.3390/land11112040.

Pathak, A. A., and B. M. Dodamani, 2020: Trend analysis of rainfall, rainy days and drought: A case study of Ghataprabha River basin, India. *Model. Earth Syst. Environ.*, **6**, 1357–1372, https://doi.org/10.1007/s40808-020-00798-7.

Pazhanivelan, S., V. Geethalakshmi, V. Samykannu, R. Kumaraperumal, M. Kancheti, R. Kaliaperumal, M. Raju, and M. K. Yadav, 2023: Evaluation of SPI and rainfall departure based on multi-satellite precipitation products for meteorological drought monitoring in Tamil Nadu. *Water*, **15**, 1435, https://doi.org/10.3390/w15071435.

Quan, H., D. Srinivasan, and A. Khosravi, 2014: Particle swarm optimization for construction of neural network-based prediction intervals. *Neurocomputing*, **127**, 172–180, https://doi.org/10.1016/j.neucom.2013.08.020.

Sahbeni, G., J. B. Pleynet, and K. Jarocki, 2023: A spatiotemporal analysis of precipitation anomalies using rainfall Gini index between 1980 and 2022. *Atmos. Sci. Lett.*, **24**, e1161, https://doi.org/10.1002/asl.1161.

Singh, A., S. Kushwaha, M. Alarfaj, and M. Singh, 2022: Comprehensive overview of backpropagation algorithm for digital

image denoising. *Electronics*, **11**, 1590, https://doi.org/10.3390/electronics11101590.

Srivastav, R. K., K. P. Sudheer, and I. Chaubey, 2007: A simplified approach to quantifying predictive and parametric uncertainty in artificial neural network hydrologic models. *Water Resour. Res.*, **43**, W10407, https://doi.org/10.1029/2006WR005352.

Sundararajan, D., 2015: *Discretewavelet Transform: A Signal Processing Approach*. John Wiley and Sons, 344 pp., https://doi.org/10.1002/9781119113119.

Tan, Y. X., J. L. Ng, and Y. F. Huang, 2023: Spatiotemporal variability assessment and accuracy evaluation of Standardized Precipitation Index and Standardized Precipitation Evapotranspiration Index in Malaysia. *Earth Sci. Inf.*, **16**, 67–89, https://doi.org/10.1007/s12145-022-00921-5.

Tanty, R., and T. S. Desmukh, 2015: Application of artificial neural network in hydrology – A review. *Int. J. Eng. Res. Technol.*, **4**, 184–188, https://doi.org/10.17577/IJERTV4IS060247.

Taormina, R., and K. W. Chau, 2015: ANN-based interval forecasting of stream flow discharges using the LUBE method and MOFIPS. *Eng. Appl. Artif. Intell.*, **45**, 429–440, https://doi.org/10.1016/j.engappai.2015.07.019.

Wang, J., Y. Gao, and X. Chen, 2018: A novel hybrid interval prediction approach based on modified lower upper bound estimation in combination with Multi-objective Salp Swarm Algorithm for short-term load forecasting. *Energies*, **11**, 1561, https://doi.org/10.3390/en11061561.

Wang, W., J. Wang, and R. Romanowicz, 2021: Uncertainty in SPI calculation and its impact on drought assessment in different climate regions over China. *J. Hydrometeor.*, **22**, 1369–1383, https://doi.org/10.1175/JHM-D-20-0256.1.

WMO, 2012: Standardized Precipitation Index user guide. WMO-1090, 16 pp., https://library.wmo.int/records/item/39629-standardized-precipitation-index-user-guide.

Zhang, G., Y. Wu, K. P. Wong, Z. Xu, Z. Y. Dong, and H. H. C. Iu, 2015: An advanced approach for construction of optimal wind power prediction intervals. *IEEE Trans. Power Syst.*, **30**, 2706–2715, https://doi.org/10.1109/TPWRS.2014.2363873.

Zhou, J., and Coauthors, 2023: Spatio-temporal changes and influencing factors of meteorological dry-wet in northern China during 1960–2019. *Sustainability*, **15**, 1499, https://doi.org/10.3390/su15021499.

Zweiri, Y. H., J. F. Whidborne, and L. D. Seneviratne, 2003: A three-term backpropagation algorithm. *Neurocomputing*, **15**, 305–318, https://doi.org/10.1016/S0925-2312(02)00569-6.