Combination of Hybrid Artificial Neural Networks with Particle Swarm Optimization Algorithm for SPEI Forecasting

MICHALA JAKUBCOVÁ,[a,b,c] PETR MÁCA,[b,c] MARTIN HANEL,[b] AND PAVEL PECH[b]

[a] *Institute of Thermomechanics of the Czech Academy of Sciences, Prague, Czech Republic*
[b] *Department of Water Resources and Environmental Modeling, Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Prague, Czech Republic*
[c] *Institute of Environmental Studies Karlovy Vary, Czech University of Life Sciences, Prague, Czech Republic*

(Manuscript received 28 February 2025, in final form 11 September 2025, accepted 18 September 2025)

ABSTRACT: Forecasting drought is critical to mitigate its potential impacts on agriculture, water resources, and ecosystems. In this study, we focused on predicting the standardized precipitation evapotranspiration index (SPEI), a widely used climatic water balance indicator. We developed a hybrid modeling framework that combines artificial neural networks (ANNs) with particle swarm optimization (PSO) to train network weights. The study evaluated the influence of four factors on SPEI forecasting performance: the PSO variant, the number of input variables, the number of hidden-layer neurons, and the choice of objective function. A total of 150 model configurations were tested using long-term meteorological data from eight U.S. catchments from the Model Parameter Estimation Experiment (MOPEX) database. Results showed that the APartPSO variant achieved the best optimization performance, and the Nash–Sutcliffe efficiency was the most effective objective function. These findings confirm that the integration of ANN with PSO is suitable for forecasting the SPEI and suggest that advanced PSO variants can be effectively applied to other inverse modeling problems in hydrology.

SIGNIFICANCE STATEMENT: Drought prediction is essential for effective water resource management and minimizing environmental and societal impacts. This study presents a new approach for forecasting the standardized precipitation evapotranspiration index by combining artificial neural networks (ANNs) with particle swarm optimization (PSO). Applied to eight U.S. catchments, the hybrid ANN–PSO model performed best when using the APartPSO variant for training and the Nash–Sutcliffe efficiency as the objective function. These results support the use of hybrid PSO–ANN models in drought monitoring and planning. The approach offers a flexible framework that can be adapted to other regions. Future research could incorporate additional climate variables to further enhance prediction accuracy.

KEYWORDS: Hydrometeorology; Neural networks; Optimization

1. Introduction

There is no universal definition of drought. In general, drought is a climatic event resulting in a prolonged deficit of water availability that affects numerous sectors (Riebsame et al. 1991; Zhaoqiang et al. 2024). The most severe impacts are often observed in agriculture, water management, energy production, ecosystems, and the economy (Mpelasoka et al. 2008; Pereira et al. 2009; Fu et al. 2024; Vadez et al. 2024; Yang et al. 2024).

There exist several types of droughts. Meteorological drought is typically defined as a period of below-average precipitation and a general lack of atmospheric moisture. Agricultural drought occurs when a precipitation deficit reduces soil moisture, thereby primarily affecting crops. Hydrological drought is associated with deficiencies in surface water and groundwater storage. When meteorological or hydrological shortfalls begin to affect drinking water supplies, hydroelectric power generation, or food production, the drought is classified as a socioeconomic drought (Wilhite and Glantz 1985; Hanel et al. 2018).

Drought severity is directly related to its impacts (Hayes et al. 2011). It is often quantified using drought indices. They are usually represented in the form of time series, and in many cases, they are based on actual measured meteorological data. One of the first comprehensive drought indices developed was the Palmer drought severity index (PDSI). It is used for analyses of the drought based on water balance equation, and it classifies the weather conditions from extremely wet to extreme drought (Palmer 1965). A simpler drought indicator is the standardized precipitation index (SPI), which is calculated based on the probability of precipitation at a location (McKee et al. 1993). A more recent and widely used metric is the standardized precipitation evapotranspiration index (SPEI), which integrates temperature through potential evapotranspiration. Due to its sensitivity to climate variability, SPEI is valuable for drought monitoring under changing climatic conditions (Haslinger et al. 2014; Hernandez and Uddameri 2014; Mossad and Alazba 2015; Karbasi et al. 2022; Sun et al. 2024).

Forecasting drought indices like SPEI is critical for proactive water management and disaster preparedness. Since physical modeling often struggles with data limitations and uncertainty, data-driven methods such as artificial neural networks (ANNs) have emerged as powerful tools for hydrological forecasting. ANNs have been effectively applied to SPI and SPEI prediction worldwide, including in Ethiopia (Belayneh and Adamowski 2012), Australia (Deo and Şahin 2015), and Iran (Hosseini-Moghari and Araghinejad 2015).

Denotes content that is immediately available upon publication as open access.

Corresponding author: Michala Jakubcová, michala.jakubcova@it.cas.cz

DOI: 10.1175/JHM-D-25-0034.1

Recent studies underscore their growing efficacy: Poudel et al. (2024) reported determination coefficient (R^2) values exceeding 0.90 for SPEI predictions across multiple temporal scales, and Esquivel-Saenz et al. (2024) achieved R^2 between 0.83 and 0.99 and root-mean-square error (RMSE) between 0.10 and 0.40 for SPEI forecasting in Mexico.

Hybrid or integrated ANN models further enhance predictive performance by combining multiple ANN structures or incorporating external optimization algorithms (Nourani and Kalantari 2010; Huo et al. 2012; Mohammadpour et al. 2018; Wable et al. 2023). In addition to single and hybrid models, stacked ensemble models have gained popularity in recent hydrological forecasting applications (Dikshit et al. 2021; Granata et al. 2022; Singh et al. 2023; Elbeltagi et al. 2024). These models combine predictions from multiple base learners, often using a metalearner, to capture nonlinear patterns and reduce prediction error. While ensemble techniques provide strong generalization capabilities, they also increase model complexity and computational cost. In contrast, this presented study focuses on optimizing a single hybrid ANN model using various particle swarm optimization (PSO) strategies, aiming to achieve competitive performance with reduced architectural and computational cost.

Traditional training of ANN weights typically relies on the backpropagation (BP) algorithm (Rumelhart et al. 1988). While widely used, BP suffers from several limitations including slow convergence and entrapment in local minima due to its reliance on gradient descent. In contrast, PSO is a population-based global search algorithm that has proven to be an effective alternative for ANN training. By simulating a swarm of particles exploring the solution space, PSO adjusts ANN weights based on individual and collective learning experiences, enabling faster convergence and avoidance of local optima (Eberhart et al. 1996; Kennedy et al. 2001; Chau 2006; Kisi et al. 2017; Ehteram et al. 2024).

Studies comparing PSO-trained ANN models to BP-trained models consistently show improved performance in terms of both accuracy and training time. For example, Chau (2007) applied a PSO–ANN model for river water level prediction and achieved better results than BP-trained models. Similarly, He et al. (2015) reported Nash–Sutcliffe (NS) efficiency values as high as 0.93 for rainfall prediction using the PSO–ANN model, while Wu et al. (2015) observed RMSE values below 70 mm.

The PSO framework is also adaptable; its performance can be further enhanced using advanced variants that modify inertia weights, neighborhood topologies, or convergence dynamics (May 1976; Máca and Pech 2015; Al-Juboori 2023). Furthermore, advanced variants of PSO, such as those using adaptive inertia weight strategies, chaotic maps, or multiswarm topologies, have shown enhanced performance in avoiding premature convergence and maintaining population diversity (Máca and Pech 2015, 2016). These enhancements are particularly relevant for drought forecasting, where high variability and nonlinearity are frequent.

Despite increasing interest in PSO–ANN models for hydrological forecasting, their application to drought index prediction remains limited. Most prior work has focused on rainfall–runoff modeling or streamflow prediction. Very few studies (Máca and Pech 2016; Al-Juboori 2023) have investigated how different PSO configurations, such as adaptive inertia strategies or multiobjective optimization, might affect drought index prediction.

This study addresses these gaps by systematically evaluating the performance of five PSO variants integrated with ANN models for SPEI forecasting across eight U.S. catchments. The novelty lies in the comparative analysis of PSO variants, variations in model architecture (number of inputs and hidden neurons), and assessment of five different objective functions. By applying these models to a large and diverse dataset, we demonstrate the performance of PSO-based training and identify the best-performing model configuration for reliable drought forecasting.

The main scientific question addressed in this study is how do different PSO variants, objective functions, and ANN configurations influence the accuracy and efficiency of SPEI drought index forecasting? Specifically, we investigate which combinations of PSO variant, input variables, number of hidden neurons, and objective function yield the most reliable predictions across diverse hydrological settings. The case study is based on data from eight U.S. catchments with 54 years of daily meteorological observations.

The remainder of this paper is organized as follows. Section 2 explains the methodology with respect to the analyzed dataset, SPEI calculations, and the ANN model configurations. Section 3 summarizes the results, which are discussed in section 4. The main findings are concluded in section 5.

2. Methodology

To represent the SPEI drought index in real-time predictive model, we applied a combination of hybrid ANN (hANN) models with PSO algorithm for forecasting SPEI. The model configurations varied in four key aspects—the number of input variables (N_{in}), the number of neurons in the hidden layer (N_{hd}), the PSO variant used for training, and the optimized objective function (OOF).

A schematic overview of the methodological framework is presented in Fig. 1, which summarizes the data flow, model configuration process, training and testing phases, and performance evaluation. The steps shown in this diagram are described in detail in the following subsections.

a. Dataset

For the hANN simulations, we used datasets from eight U.S. catchments. The meteorological data were obtained from the Model Parameter Estimation Experiment (MOPEX) project, which serves as a benchmarking dataset of hydrological models (Duan et al. 2006).

Table 1 lists the key characteristics of each catchment, including latitude (lat), longitude (lon), drainage area (area), monthly mean precipitation (mP), and monthly mean evapotranspiration (mET) over the observation period.

We analyzed the dataset from 1948 to 2002 using a split-sample test approach. The data were divided into a calibration period (1948–75) and a validation period (1976–2002) to ensure independent assessment of model performance. The daily records of precipitation and evapotranspiration from the

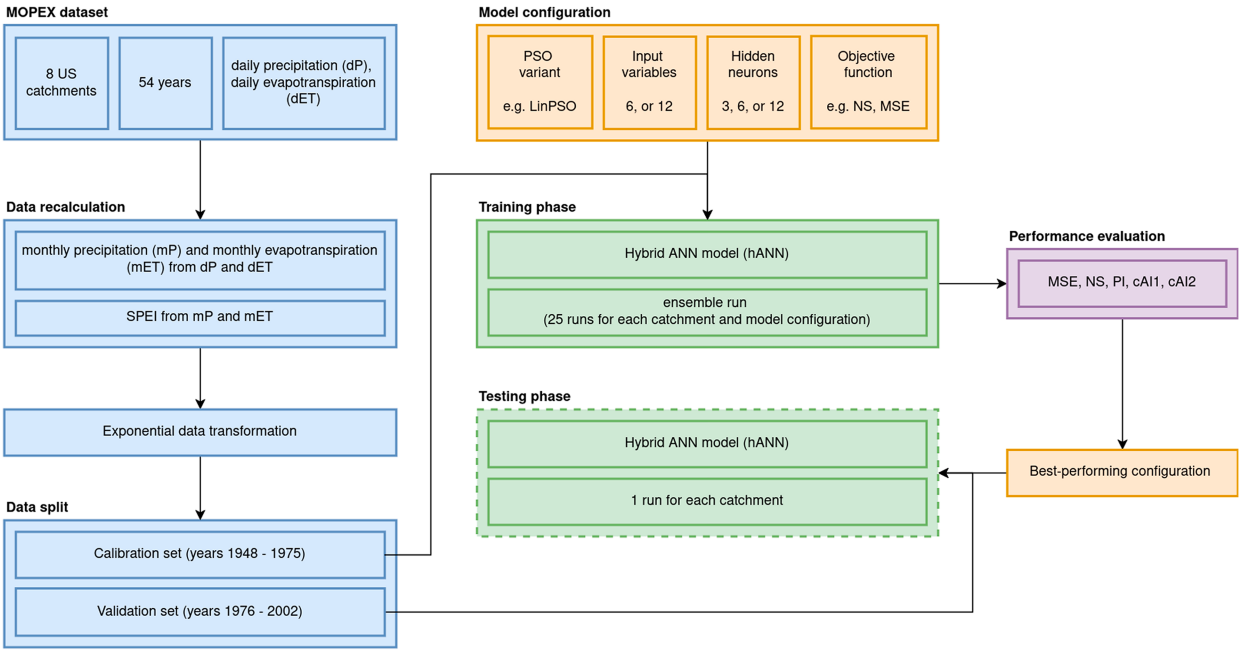

FIG. 1. Schematic representation of the modeling framework used for SPEI forecasting with hANN–PSO models. The workflow includes data preparation (blue), model configuration (orange), training and testing phases (green), and performance evaluation (purple).

MOPEX database were aggregated to monthly values, from which the SPEI series were calculated.

We applied a nonlinear transformation on the original data D_o (i.e., SPEI and SPI values) before using them in the ANN model. The transformed data D_t are calculated as

$$D_t = 1 - \exp(-0.015 D_o). \tag{1}$$

All values of D_t are thus normalized for the input to the ANN model. Before calculating the accuracy criteria during training of the ANN, the model outputs were transformed back to the original scale using the inverse of that transformation as

$$D_o = \begin{cases} \dfrac{1}{0.015}\ln(10E + 30), & \text{if} \quad (1 - D_t) = 0 \\ \dfrac{1}{0.015}\ln\left(\dfrac{1}{1 - D_t}\right), & \text{if} \quad (1 - D_t) \neq 0\,. \end{cases} \tag{2}$$

This exponential transformation was applied to normalize the data range and reduce the influence of extreme values, which can impair ANN training. Similar strategies have been used in previous hydrological modeling studies to improve training stability and generalization and reduce sensitivity to outliers (Máca et al. 2014; Máca and Pech 2016). Equation (1) describes the nonlinear transformation applied to the original input data, using an exponential decay function to scale all values to a normalized range suitable for input into the ANN. This transformation helps in stabilizing the training process by reducing the influence of large input magnitudes. Equation (2) is the inverse of this transformation, converting the normalized ANN output back into the original data scale. This step is essential for comparing model outputs to observed data in their natural units during performance evaluation.

Although SPEI and SPI are standardized indices, individual time series can still contain occasional extreme values that have an outsized influence on model training and error calculations. The applied transform ensures that such extremes remain within a narrow, bounded range, promoting consistent convergence behavior. Since the transformation is strictly monotonic, all temporal relationships between inputs and targets are preserved. Predictions are always back-transformed to the original scale for evaluation, thus retaining interpretability.

Table 2 shows that the raw SPEI inputs and targets were already centered near zero, with standard deviations close to one, negligible skewness, and only slightly sub-Gaussian tails (kurtosis approximately 2.4). Applying the exponential transform preserved the symmetry (skewness) and tail weight (kurtosis) of the distribution but substantially compressed its dynamic range. This uniform range compression reduces the numerical leverage of rare extremes while maintaining their rank order, which can improve the numerical stability of ANN training by reducing gradient magnitudes and mitigating the influence of large outliers during weight updates.

TABLE 1. Characteristics of selected catchments.

USGS ID	Lat (°)	Lon (°)	Area (km^2)	mP (mm)	mET (mm)
01127000	41.5980	−71.9850	1147	99.88	60.20
01197500	42.2320	−73.3550	454	99.92	55.03
01321000	43.3528	−74.2708	790	103.33	57.49
01371500	41.6860	−74.1660	1144	94.71	63.10
01372500	41.6531	−73.8731	291	86.83	58.40
01426500	42.0031	−75.3839	957	91.28	59.01
01445500	40.8306	−74.9786	171	104.30	65.72
01503000	42.0353	−75.8033	3591	85.91	59.24

TABLE 2. Aggregated descriptive statistics of SPEI inputs and targets across all catchments, before and after applying the exponential transformation [Eq. (1)]. The statistics include mean, standard deviation (std dev), skewness, kurtosis (Pearson's definition), and minimum and maximum values.

	Raw inputs	Transformed inputs	Raw targets	Transformed targets
Mean	−0.0087	−0.0002	−0.0018	−0.0001
Std dev	0.9847	0.0148	0.9910	0.0149
Skewness	0.0021	−0.0289	0.0016	−0.0292
Kurtosis	2.4000	2.4044	2.3849	2.3897
Min	−2.4234	−0.0370	−2.4234	−0.0370
Max	2.3496	0.0346	2.3496	0.0346

b. SPEI

We applied the ANNs to forecast SPEI. The input data for SPEI calculation are monthly time series of climatic water balance, where the potential evapotranspiration is subtracted from precipitation (Vicente-Serrano et al. 2010).

SPEI is a variant of SPI that accounts for the influence of potential evapotranspiration (Vicente-Serrano et al. 2010). It is computed by determining the cumulative probability of the observed climatic water balance at a location. The probabilities are then converted into a standard normal distribution, which creates the final value of the SPEI (McKee et al. 1993; Stagge et al. 2015).

In this study, we computed SPEI at a monthly time scale with a 12-month accumulation period, using an unshifted rectangular kernel for aggregation. We assumed a log-logistic probability distribution for the water balance, with parameters fitted by unbiased probability-weighted moments (Beguería et al. 2014). The SPEI values were calculated using the SPEI package in the R programming language (Beguería and Vicente-Serrano 2013).

c. ANN Models

We employed a multilayer perceptron (MLP) ANN architecture, which is a standard suitable universal approximator (Hornik et al. 1989). The MLP used in this study consists of one input layer, one hidden layer of neurons, and one output layer with one output neuron. The network topology is fully connected and transfer of information is feed-forward.

In our models, we used processing elements (i.e., components of ANN where computations are performed) with external biases (Eberhart and Shi 2007). The normalized transformed simulated output MOD_t from the ANN model at a given time interval is calculated as

$$\mathrm{MOD}_t = w_b + \sum_{j=1}^{N_{\mathrm{hd}}} w_j f(a), \tag{3}$$

where w_b is the bias weight into the output neuron, w_j is the weight of jth hidden neuron into the output neuron, $f()$ is the activation function, and a is the neuron's activation.

We chose the RootSig activation function for all neurons. RootSig produces bounded outputs like a sigmoid but with a smoother gradient near the origin, which helps mitigate vanishing gradient issues while maintaining stability compared to unbounded functions such as rectified linear unit (ReLU). Previous comparative studies in both financial time series forecasting and hydrological modeling have reported that RootSig achieves superior predictive skill compared to classical functions like tanh or logistic sigmoid (Duch and Jankowski 1999; da S. Gomes et al. 2011). Notably, Máca et al. (2014) demonstrated that, among eleven commonly used activation functions, RootSig provided the best overall performance in flood runoff forecasting. Given the parallels between runoff and drought index prediction as nonlinear time series problems, we adopted RootSig as the activation function in this study.

The RootSig activation function is, in general,

$$f(a) = \frac{a}{1 + \sqrt{1 + a^2}}, \tag{4}$$

with the activation defined as

$$a = w_{bj} + \sum_{i=1}^{N_{\mathrm{in}}} w_{ij}\mathrm{OBS}_{ti}, \tag{5}$$

where w_{bj} is weight of the bias neuron entering the jth neuron in the hidden layer, w_{ij} is weight of the ith input entering the jth neuron in the hidden layer, and OBS_{ti} is the ith transformed input into the network.

We trained the ANN weights using the PSO algorithm. Five different statistical metrics were used as objective functions (optimization criteria) during training, as is common in hydrological modeling. For the SPEI simulations, we employed integrated ANN models with various settings. The PSO methods, OOF, and integrated ANN model configuration are explained in the next section of the paper.

The training algorithms were implemented in C++ and executed on a 64-bit Linux system. Most simulations were carried out with the support of the MetaCentrum computing infrastructure, provided under the Czech Education and Scientific Network (CESNET) project. All postprocessing calculations were performed using the R statistical software environment (R C Team 2021) and Python programming language.

1) PSO VARIANTS FOR ANN TRAINING

We used PSO to train the ANN models. PSO is a population-based optimization algorithm that iteratively moves a swarm of candidate solutions (particles) through the parameter space. Each particle updates its position following the rules introduced by Kennedy and Eberhart (1995) to collectively search for an optimal solution.

Particle's position $\mathbf{X}_i^{t+1}$ in iteration $t + 1$ in the space is calculated as

$$\mathbf{X}_i^{t+1} = \mathbf{X}_i^t + \mathbf{V}_i^{t+1}. \tag{6}$$

Each position in the population represents a vector of searched parameters. With respect to the ANN model, the parameters are the weights of the ANN model.

The velocity of particles $\mathbf{V}_i^{t+1}$ is updated for all $i = 1, \ldots, S$, where S is the total number of particles in the swarm population. The velocity equation is

$$\mathbf{V}_i^{t+1} = W \cdot \mathbf{V}_i^t + c_1 \cdot \mathbf{U}_1^t \otimes (\mathbf{P}_i^t - \mathbf{X}_i^t) + c_2 \cdot \mathbf{U}_2^t \otimes (\mathbf{G}^t - \mathbf{X}_i^t), \tag{7}$$

where W is the parameter of inertia weight, c_1 and c_2 denote the acceleration constants predefined by the user, $\mathbf{U}_1$ and $\mathbf{U}_2$ are the independent random vectors sampled from a uniform distribution in the range [0, 1], $\mathbf{P}_i = (p_{i1}, p_{i2}, \ldots, p_{i\mathrm{Dim}})$ is the best particle's position achieved so far representing the cognition knowledge of swarm particles, and $\mathbf{G} = (g_1, g_2, \ldots, g_{\mathrm{Dim}})$ is the best location of all particles achieved so far controlling the social influence of swarm particles (Kennedy and Eberhart 1995; Simon 2013).

The methodological novelty of this study lies in the comparative evaluation and integration of five distinct PSO modifications with multiple objective functions. These PSO variants, including PSO method with linearly decreasing inertia weight (LinPSO), PSO method with random chaotic model (ChaoPSO), PSO method with nonlinearly decreasing inertia weight (NonlinPSO), adaptive PSO (AdaptPSO), and adaptive multi-swarm PSO with particle redistribution (APartPSO), were used to optimize ANN model weights. The effect of inertia weight dynamics, adaptive behavior, and swarm coordination are central to improving convergence in high-dimensional inverse problems such as drought index forecasting.

All PSO versions update the particle's velocity using a parameter of inertia weight, and we compared all these versions in our previous work (Jakubcová et al. 2014, 2015; Máca and Pech 2015). The used PSO methods with the corresponding formulas for calculating the inertia weight are listed in Table 3.

The LinPSO was first used by Shi and Eberhart (1999). The parameter of W decreases linearly during iterations from $W_{\max} = 0.9$ to $W_{\min} = 0.4$. The ChaoPSO was used by Feng et al. (2007). The inertia weight changes during iterations based on a random number $\mathbf{U}^t$ from the range [0, 1] and auxiliary variable $z = 4z(1 - z)$, where the initial value of z is uniformly distributed in [0, 1]. The NonlinPSO uses parameter u, which is set to 1.0002, and initial value of inertia weight W_{ini} uniformly distributed in [0, 1] (Nickabadi et al. 2011).

Moreover, two adaptive strategies of inertia weight were used. First, the AdaptPSO proposed by Nickabadi et al. (2011) calculates the inertia weight through success percentage of the swarm ($\mathbf{Ps}^t$). It is calculated as

$$\mathbf{Ps}^t = \frac{\sum_{i=1}^{n} \mathbf{S}_i^t}{n}, \tag{8}$$

where n is the size of the population and $\mathbf{S}_i$ shows the success of ith particle, which is equal to 1 if $f(\mathbf{P}_i^t) < f(\mathbf{P}_i^{t-1})$ or equal to 0 if $f(\mathbf{P}_i^t) = f(\mathbf{P}_i^{t-1})$. In this approach, $W_{\min} = 0$ and $W_{\max} = 1$.

Second adaptive strategy is APartPSO modification proposed by Jakubcová et al. (2015). It combines the global exploration and local exploitation in the space according to the development of particle's location. If the particle improves its position compared to its previous locations, the 1) equation from Table 3 is used; otherwise, the 2) equation is applied. In this variant, $W_{\min} = 0.1$ and $W_{\max} = 0.9$. In particular, the use of APartPSO, which balances local exploitation and global exploration through performance-sensitive inertia adaptation, represents an advanced optimization approach not previously applied to SPEI prediction.

TABLE 3. Applied PSO modifications for training ANN models.

PSO	Inertia weight
LinPSO	$\mathbf{W}_i^t = \frac{t_{\max} - t}{t_{\max}}(W_{\max} - W_{\min}) + W_{\min}$
ChaoPSO	$\mathbf{W}_i^t = 0.5\mathbf{U}^t + 0.5z$
NonlinPSO	$\mathbf{W}_i^t = W_{\mathrm{ini}} u^t$
AdaptPSO	$\mathbf{W}_i^t = (W_{\max} - W_{\min})\,\mathbf{Ps}^t + W_{\min}$
APartPSO	1) $\mathbf{W}_i^t = \left(\frac{W_{\max} + W_{\min}}{2} - W_{\min}\right)\mathbf{U}^t + W_{\min}$
	2) $\mathbf{W}_i^t = \left(\frac{W_{\max} + W_{\min}}{2} - W_{\min}\right)\mathbf{U}^t + \left(\frac{W_{\max} + W_{\min}}{2}\right)$

The PSO settings used in this study were based on distributed versions of PSO (Yan et al. 2007; Jakubcová et al. 2014, 2015), with three complexes, 40 particles per complex, and 20 generations per complex. Shuffling and redistribution of particles across complexes were performed 10 times. Each optimization run terminated when the maximum number of function evaluations was reached, and a total of 25 independent runs were conducted. The acceleration coefficients were set to $c_1 = c_2 = 2$, following the canonical PSO formulation of Eberhart and Shi (2000), which has been widely adopted as a robust default. Swarm sizes in the range of 20–40 particles and generation counts on the order of tens to a few hundred are commonly recommended as a trade-off between convergence reliability and computational cost (Poli et al. 2007; Zambrano-Bigiarini and Rojas 2013). Thus, our configuration reflects prior successful applications of PSO in hydrological and environmental modeling, ensuring comparability with established practice.

2) OBJECTIVE FUNCTIONS

The OOFs are in hydrological modeling commonly used accuracy criteria (Dawson et al. 2007; Bennett et al. 2013). The evaluated criteria are mean square error (MSE), NS, persistence index (PI), and two combined accuracy indices (cAI1 and cAI2). The cAI1 and cAI2 are composed of two different objective functions, both of which influence the final criterion with a given weight.

The formulas of analyzed objective functions are defined below, where OBS is the observed SPEI, MOD is the modeled SPEI, $\overline{\mathrm{OBS}}$ is the mean of the observed SPEI, and n is the total number of observations. Consider

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{OBS}[i] - \mathrm{MOD}[i])^2, \tag{9}$$

$$\mathrm{NS} = 1 - \frac{\sum_{i=1}^{n}(\mathrm{OBS}[i] - \mathrm{MOD}[i])^2}{\sum_{i=1}^{n}(\mathrm{OBS}[i] - \overline{\mathrm{OBS}})^2}, \tag{10}$$

$$\mathrm{PI} = 1 - \mathrm{tPI}, \tag{11}$$

$$\mathrm{cAI1} = 0.85\mathrm{tPI} + 0.15\mathrm{MSE}, \tag{12}$$

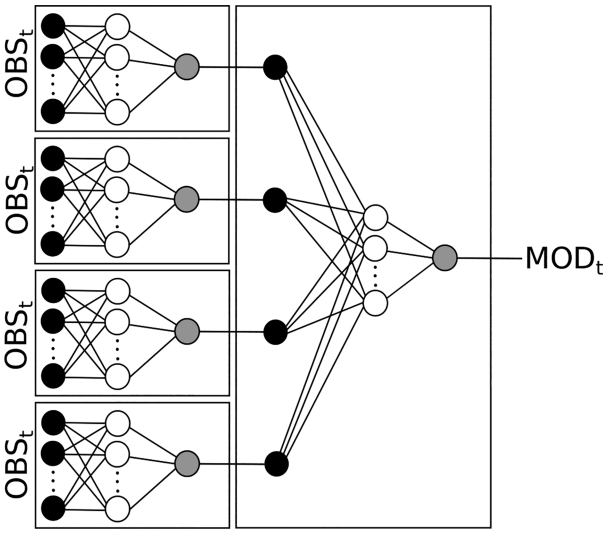

FIG. 2. Integrated ANN models into hANN. Circles filled with black represent input layer, circles filled with white represent hidden layer, and circles filled with gray represent outputs. The fifth ANN was trained exclusively on calibration outputs of the four base ANNs, ensuring that no information from the validation period was used.

$$\text{cAI2} = 0.85\text{MAE} + 0.15\text{dRMSE}, \tag{13}$$

where tPI is the transformed persistence index, MAE is the mean absolute error, and dRMSE is the RMSE in derivatives. The equations are

$$\text{tPI} = \frac{\sum_{i=1}^{n} (\text{OBS}[i] - \text{MOD}[i])^2}{\sum_{i=1}^{n} (\text{OBS}[i] - \text{OBS}[i-1])^2}, \tag{14}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\text{OBS}[i] - \text{MOD}[i]|, \tag{15}$$

$$\text{dRMSE} = \sqrt[4]{\frac{1}{n-1}\sum_{i=1}^{n-1} [(\text{OBS}[i] - \text{OBS}[i+1]) - (\text{MOD}[i] - \text{MOD}[i+1])]^4}. \tag{16}$$

All abovementioned OOFs were used not only for ANN training during calibration of the models but also for result analyses. The objective functions serve as accuracy criteria for evaluation of the models' performance.

In addition to continuous skill metrics [e.g., NS efficiency (NSE)], we also evaluated categorical drought detection skill using thresholds (SPEI $\leq -1.0, -1.5, -2.0$), expressed in terms of hit rate, false alarm ratio, and Heidke skill score. Details and plots are provided in the appendix.

3) INTEGRATED ANN MODELS

In our research, always five ANN models were integrated into one hANN model. The outputs from four models are inputs into the fifth model as it is displayed in Fig. 2. The fifth ANN acts as an error-correcting metalearner, trained only on calibration predictions produced by the four base ANNs. The base models are first trained on calibration data, their outputs are then used as inputs to the meta-ANN, ensuring no target information from the validation period leaks into the meta-training process. The final forecasted SPEI drought index is the output from the fifth ANN model.

The hANNs are trained in a lead-0 (nowcasting) setup, where the aim is to predict the current month's SPEI using only information available up to month $[t-1]$. Thus, the target is SPEI$[t]$, while the inputs are lagged values SPEI$[t-1]$, SPEI$[t-2]$, etc., corresponding to a 0-month prediction horizon.

The hANN models differ in N_{in}, N_{hd}, and PSO method used for training and OOF. The N_{in} is equal to 6 or 12 with the input variables according to Table 4, where SPI is obtained from precipitation data. The N_{hd} is either 3 or 6. The PSO method used for training is the same for the whole hANN, and it is one of the five variants mentioned above in the text.

The optimized objective criteria are different for each of the ANN model within the hANN. It is the main advantage of the hANN because the performance of each model is combined, and the fifth ANN model works as a correction model (Goswami et al. 2005; Huo et al. 2012).

The general form of the hANN model within this paper is N_{in}-N_{hd}-PSO-OOF, e.g., 6-3-LinPSO-MSE. The OOF in the notation is the objective function optimized by the final fifth ANN model. To distinguish between input variable SPEI and SPI for $N_{\text{in}} = 12$, we used $N_{\text{in}} = 12$ for SPEI, and $N_{\text{in}} = 12s$ for SPI in the notation of the hANN. The PSO and OOF in the notation in the validation period indicate the optimization method and objective function used for obtaining model weights during calibration.

TABLE 4. Input variables into the ANN models for each number of inputs.

n_{in}	Input variable
6	SPEI$[t-c]$, for $c = 1, 2, \ldots, 6$
12	SPEI$[t-c]$, for $c = 1, 2, \ldots, 12$
12	SPI$[t-c]$, for $c = 1, 2, \ldots, 12$

TABLE 5. Influence of each factor on the accuracy criteria. Results for calibration are on the left, for validation on the right side of each column. The significance level *** is for P value ≤ 0.001, ** for $0.001 < P$ value ≤ 0.01, * for $0.01 < P$ value ≤ 0.05, · for $0.05 < P$ value ≤ 0.1, and no sign for P value > 0.1.

	Accuracy criteria									
	MSE		NS		PI		cAI1		cAI2	
Factor	cal.	val.	cal.	val.	cal.	val.	cal.	val.	cal.	val.
Catch.	***	***	***	***	***	***	***	***	***	***
N_{in}	***	*		***	***	***	***	.	***	***
N_{hd}	***	***	***	***	***	***	***	***	***	
OOF	***	***			***	.	***	***	***	
PSO	.		***	**	***	***	*			

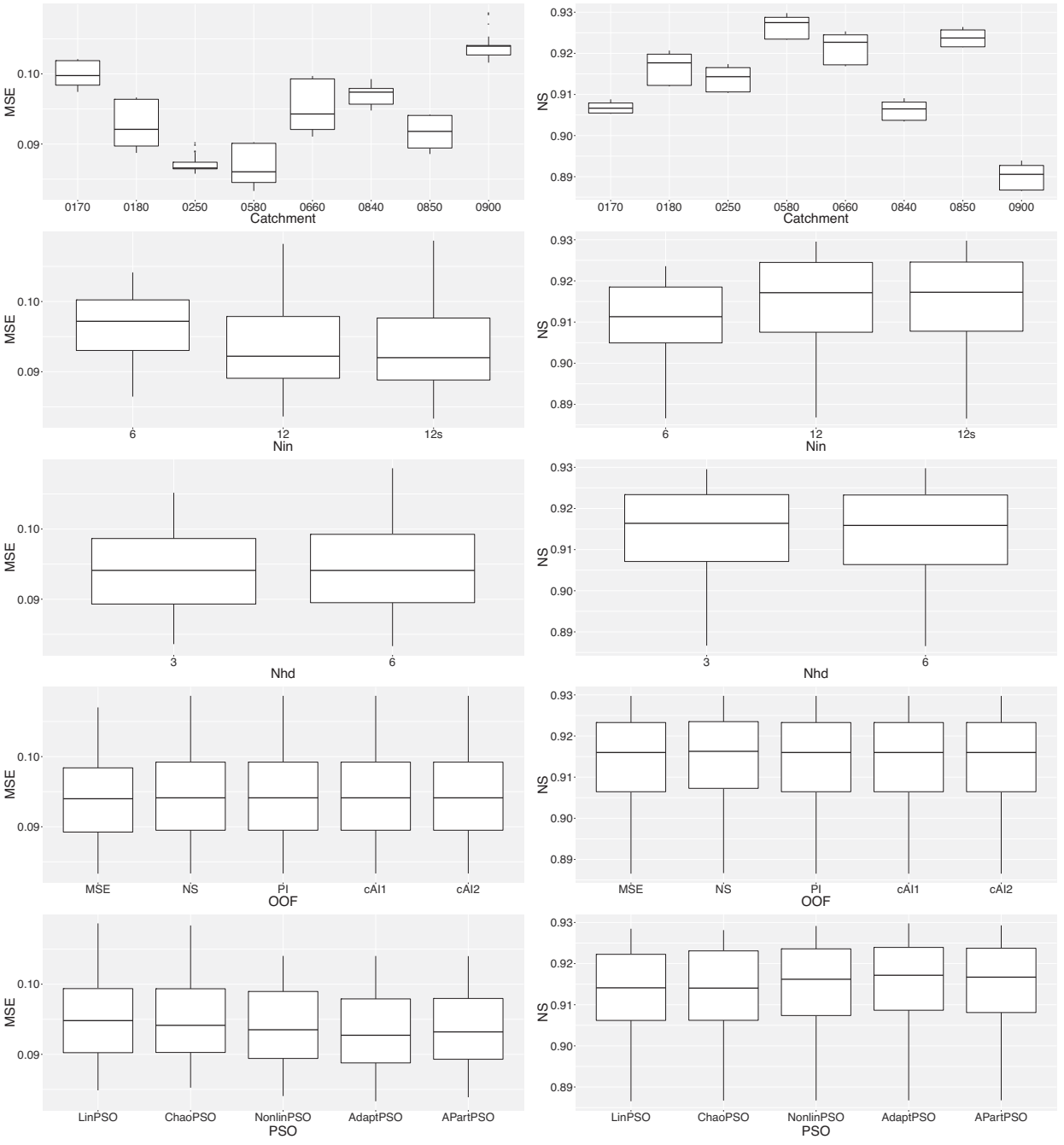

FIG. 3. Boxplots of all achieved values of (left) MSE and (right) NS (during calibration based on different factors.

The total number of hANN models for each catchment is 150 (i.e., 3 sets of inputs × 2 sets of neurons in the hidden layer × 5 PSO variants × 5 OOF).

4) BASELINE MODELS AND REGIME ANALYSIS

We compared the performance of the proposed hANN with two common benchmark models—persistence and seasonal climatology. These baselines provide complementary reference points for assessing model skill. Persistence forecasts assume that the next month's value is equal to the most recently observed value, capturing short-term autocorrelation but ignoring seasonal effects. Seasonal climatology forecasts use the long-term historical mean for the corresponding calendar month, representing purely seasonal variability without memory of recent anomalies.

When the target series is stable and exhibits high month-to-month autocorrelation, persistence is typically difficult to outperform. In contrast, when a strong seasonal cycle is present but

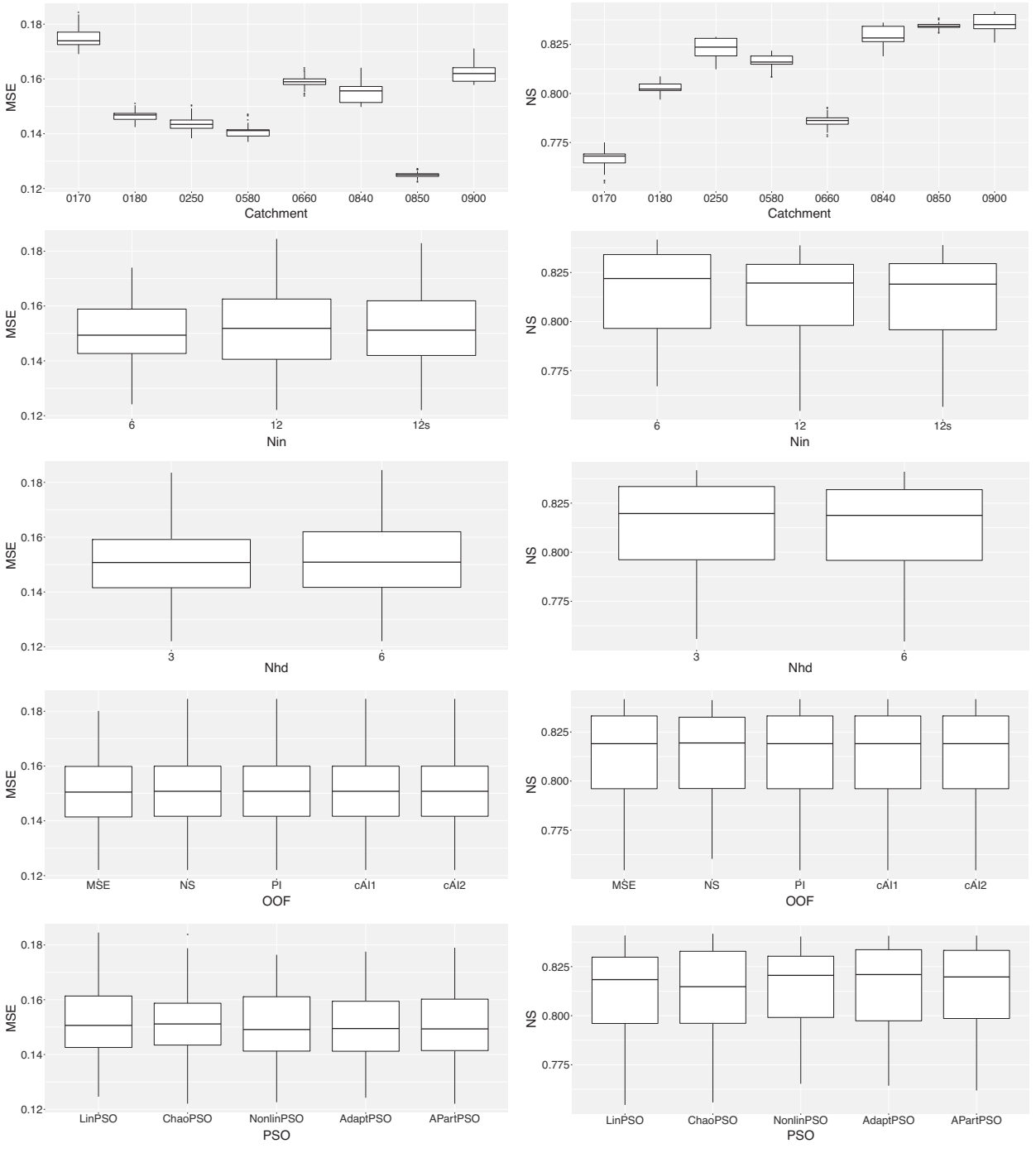

FIG. 4. Boxplots of all achieved values of (left) MSE and (right) NS during validation based on different factors.

short-term autocorrelation is weak, seasonal climatology can be competitive (Namias 1952; Tuel and Martius 2023). By including both baselines, we can better assess whether the hANN's performance gains are due to its ability to capture persistence-like memory, seasonal structure, or both.

To further investigate model behavior under different hydroclimatic conditions, the validation period was divided into two regimes based on the magnitude of month-to-month changes in the target SPEI values. For each time step t, the absolute difference from the previous step was calculated as

$$\Delta\text{SPEI}_t = |\text{SPEI}_t - \text{SPEI}_{t-1}|. \tag{17}$$

Months with $\Delta\text{SPEI}_t \geq 0.5$ were classified as transition months, representing periods of rapid hydroclimatic change. Months with $\Delta\text{SPEI}_t < 0.5$ were classified as stable months,

TABLE 6. The best levels of each factor for each accuracy criteria and the final best level based on Tukey's HSD test. A dash indicates no significant difference in levels. The asterisk (*) indicates 1 = MSE, 2 = NS, 3 = PI, 4 = cAI1, and 5 = cAI2. Two asterisks (**) indicate 1 = LinPSO, 2 = ChaoPSO, 3 = NonlinPSO, 4 = AdaptPSO, and 5 = APartPSO.

Factor	MSE	NS	PI	cAI1	cAI2	Final
			Calibration period			
Catchment	01371500	01371500	01197500	01197500	01371500	01371500
N_{in}	12, 12*s*	—	6	12, 12*s*	12, 12*s*	12, 12*s*
N_{hd}	6	6	6	6	6	6
OOF*	2, 3, 4, 5	—	1, 2, 4, 5	1, 2, 3, 5	1, 2, 3, 4	2
PSO**	—	5	5	—	—	5
			Validation period			
Catchment	01445500	01503000	01127000	01372500	01445500	01445500
N_{in}	—	6	6	—	12, 12s	6
N_{hd}	6	6	6	6	—	6
OOF*	2, 3, 4, 5	—	—	1, 2, 3, 5	—	2, 3, 5
PSO**	—	4, 5	2, 3, 4, 5	—	—	4, 5

indicating relatively constant conditions. The first month of the validation period was assigned to the stable regime by default, as no preceding observation was available for change computation. Model performance metrics were then computed separately for each catchment and each regime to identify differences in relative model skill.

3. Results

To obtain a representative sample of outcomes, each model configuration was run 25 times for the calibration period and then evaluated once on the validation period. The total number of runs was 31 200 (i.e., 3 sets of inputs × 2 sets of neurons in hidden layer × 5 PSO variants × 5 OOF × 8 catchments × 26 runs).

a. Statistical evaluation

The analyzed catchment, N_{in}, N_{hd}, OOF, and PSO variant were treated as factors potentially influencing the resulting performance metrics. The output data were analyzed using statistical tests, and the results are summarized in the tables and figures that follow.

We employed analysis of variance (ANOVA) to statistically evaluate the results. Table 5 shows the significance level for each factor from the ANOVA. If a factor is significant at $\alpha = 0.05$ (indicated by at least one star in the table), it means that at least one level of that factor produces results that differ significantly from the others. For example, all accuracy metrics showed a dependence on the choice of catchment, whereas the PSO variant had a comparatively lower effect (lower significance level).

Table 5 is supplemented by Figs. 3 and 4, which display boxplots of selected accuracy criteria. We focus on MSE and NS in these figures since they are among the most commonly used performance statistics in hydrological modeling.

From the boxplots (Figs. 3 and 4), one can assess which level of each factor yields significantly different performance compared to the others. For both calibration and validation, using $N_{hd} = 6$ and the APartPSO variant appears to give the best results, whereas using MSE as the objective function

TABLE 7. Statistical indices of the best hANN models.

	MSE	NS	PI	cAI1	cAI2
			Calibration period: 12-6-APartPSO-NS		
Min	9.33×10^{-2}	7.30×10^{-1}	-9.83×10^{-1}	8.40×10^{-1}	3.08×10^{-1}
25%	1.08×10^{-1}	8.68×10^{-1}	-3.58×10^{-1}	9.14×10^{-1}	5.32×10^{-1}
Median	1.14×10^{-1}	8.83×10^{-1}	-3.09×10^{-1}	9.99×10^{-1}	5.53×10^{-1}
75%	1.25×10^{-1}	8.91×10^{-1}	-2.55×10^{-1}	1.05×10^{0}	5.82×10^{-1}
Max	1.60×10^{-1}	9.06×10^{-1}	-8.04×10^{-2}	1.34×10^{0}	6.50×10^{-1}
Mean	1.16×10^{-1}	8.78×10^{-1}	-3.18×10^{-1}	9.95×10^{-1}	5.56×10^{-1}
Std dev	1.32×10^{-2}	1.86×10^{-2}	9.35×10^{-2}	9.71×10^{-2}	3.36×10^{-2}
			Validation period: 6-6-AdaptPSO-NS		
Min	1.30×10^{-1}	5.88×10^{-1}	-6.85×10^{-1}	8.83×10^{-1}	4.01×10^{-1}
25%	1.56×10^{-1}	7.50×10^{-1}	-2.82×10^{-1}	9.30×10^{-1}	6.67×10^{-1}
Median	1.70×10^{-1}	7.75×10^{-1}	-2.16×10^{-1}	9.89×10^{-1}	6.95×10^{-1}
75%	1.80×10^{-1}	7.95×10^{-1}	-1.64×10^{-1}	1.02×10^{0}	7.16×10^{-1}
Max	2.19×10^{-1}	8.26×10^{-1}	-2.81×10^{-2}	1.24×10^{0}	7.84×10^{-1}
Mean	1.69×10^{-1}	7.71×10^{-1}	-2.29×10^{-1}	9.89×10^{-1}	6.92×10^{-1}
Std dev	1.82×10^{-2}	3.14×10^{-2}	8.95×10^{-2}	6.86×10^{-2}	3.81×10^{-2}

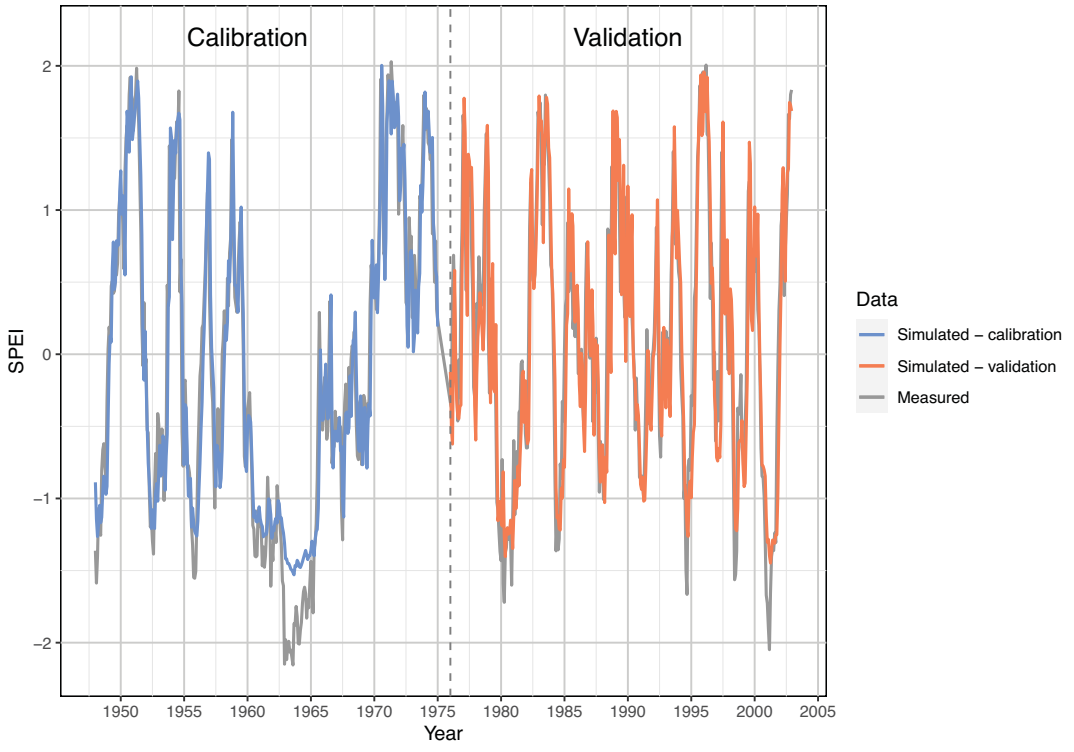

FIG. 5. Measured and simulated SPEI time series during the calibration and validation periods in catchment 01371500. Results correspond to a representative hANN model 12-6-APartPSO-NS, shown here as an illustrative example of the best-performing run.

resulted in worse performance than the other OOF choices. The optimal choice of N_{in} differs between calibration (N_{in} = 12 or 12*s*) and validation (N_{in} = 6), as does the choice of catchment.

To rigorously identify the best level of each factor across all metrics, we performed a post hoc Tukey's honest significant difference (HSD) test for multiple comparison of means (Ott and Longnecker 2008). This statistical analysis helped generalize our conclusions across all the accuracy criteria.

Table 6 summarizes the best-performing level of each factor for both calibration and validation. The results show that certain factor levels are clearly superior, whereas in some cases, there is no significant difference between two or more levels.

Based on these results, we identified the best hANN configuration. For the calibration period, two hANN models with different N_{in} showed equivalent performance. Both of these best-calibration models used 12 inputs, 6 neurons in the hidden layer optimized by NS criterion with APartPSO training method.

For the validation period, six hANN models with three OOF and two PSO factors showed statistically indistinguishable performance. The best-performing validation models all had six SPEI inputs and six neurons in the hidden layer.

b. *Best hANN models*

Each hANN configuration was trained 25 times with different PSO initializations, and the run achieving the highest NS was designated as the best hANN for illustrative comparison. We estimated two overall best hANN for calibration and six overall best hANN for validation out of the total 150 possible hANN models. For clarity, we selected one best hANN model for calibration and one for validation as the representative models. The models are 12-6-APartPSO-NS and 6-6-AdaptPSO-NS, respectively. It should be noted that our statistical conclusions are based on the distribution of all 25 runs for each configuration (as depicted in the boxplots and ANOVA/Tukey's tests), not on these single illustrative runs.

The overall performance of the hANN is explained by statistical indices of the model accuracy criteria obtained during calibration and validation. Table 7 displays the minimum, 25% quartile, median, 75% quartile, maximum, mean, and standard deviation of all achieved values for the best estimated hANN models. The results show that during calibration, lower MSE, cAI1 and cAI2, and higher NS were obtained. During validation, the PI criterion was higher but still smaller than zero, which indicates that the naive model was better than hANN in many cases. The median of MSE for both hANNs is approximately 1.42×10^{-1}, and the median of NS is 8.29×10^{-1}, which indicates a good model fit.

Figures 5 and 6 present the time series of measured and simulated SPEI drought index. Models used for the visualization are the estimated best hANN for calibration and validation. The displayed catchments are the final best according to Table 6. It is evident that the simulated SPEI is close to the measured one and that the models provide sufficient

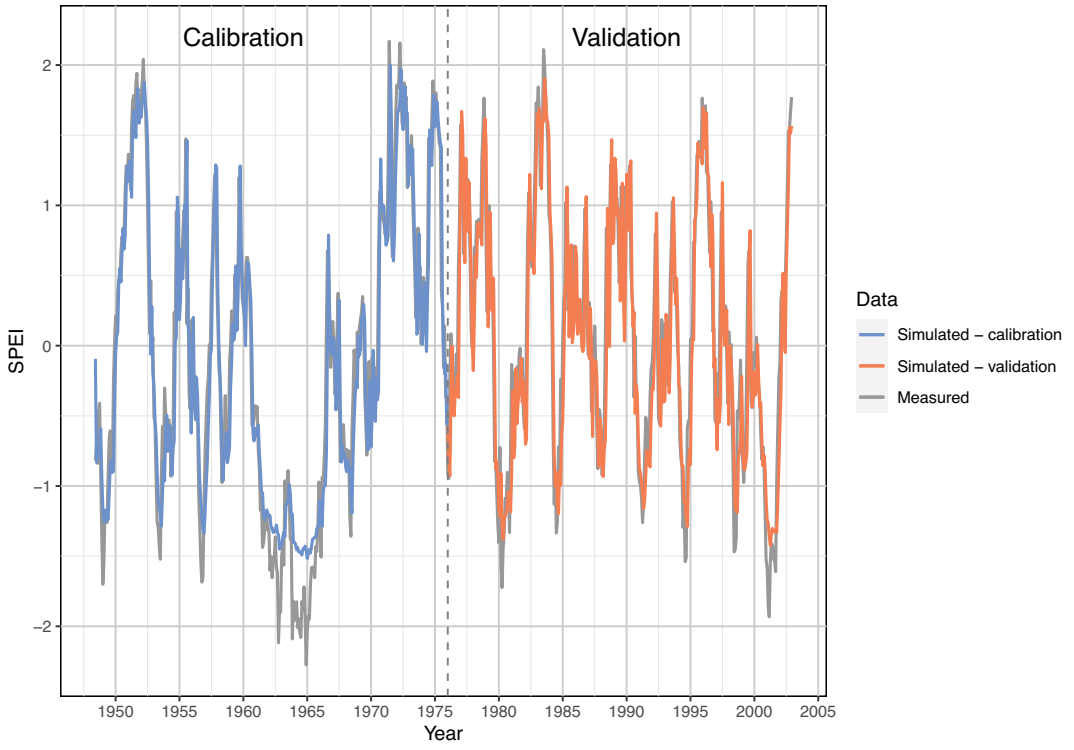

FIG. 6. Measured and simulated SPEI time series during the calibration and validation periods in catchment 01445500. Results correspond to a representative hANN model 6-6-AdaptPSO-NS, shown here as an illustrative example of the best-performing run.

forecasts. The best hANN obtained during calibration provides good fit also for validation data, and vice versa. The only problem could be the overestimation of the lower values of SPEI.

The best hANN models for each catchment are displayed in Tables 8 and 9. They were estimated based on the performance according to NS objective function, and thus, these models provided the highest NS values. No statistical evaluation was performed within this analysis.

From the tables, it is evident that during calibration, only one model for each catchment gave the best results, except for the catchment with the USGS ID 01127000. On the other hand, more hANN models provided the same NS value for each catchment during validation, except for the catchment with USGS ID 01321000. This is related to the statistical evaluation, based on which we determined that there is no significant difference between more levels of each factor during validation than during calibration.

c. *Comparison with the baseline models*

Figure 7 compares the NS for the best-performing hANN, persistence, and seasonal climatology models across all eight catchments, separated into stable and transition months in the validation period. In the stable regime, persistence and hANN achieved similarly high NS values (>0.85) for all catchments, while seasonal climatology lagged behind, often producing negative performance scores. This indicates that when conditions change slowly and short-term autocorrelation is high, the most recent observation (persistence) provides a strong predictive signal that is difficult to surpass, even for a trained ANN.

In the transition regime, NS values were generally lower for all models, reflecting the increased difficulty of predicting rapid hydroclimatic shifts. In most catchments, the hANN outperformed persistence, suggesting that the nonlinear relationships captured by the network can provide added value during periods of change. However, this advantage was not universal, and in some cases, persistence still matched or exceeded hANN skill.

To further highlight the relative benefit of hANN over the persistence benchmark, Fig. 8 shows the difference in NS ($\Delta NS = NS_{hANN} - NS_{Persistence}$) for each catchment during stable and transition months, respectively. In the stable regime, ΔNS values were generally small and negative, confirming that hANN does not systematically outperform persistence under highly autocorrelated conditions. Conversely, during transition months, ΔNS was positive for most catchments, indicating that the hANN is able to extract additional predictive information beyond simple memory effects when rapid hydroclimatic changes occur.

Seasonal climatology consistently exhibited the weakest performance in both regimes (Fig. 7), underscoring its limited utility when short-term dynamics are important.

Overall, these results demonstrate that hANN performance gains over persistence are both catchment- and regime-dependent, which explains the presence of negative PI values in certain validation cases (Tables 8 and 9).

TABLE 8. The best achieved accuracy criteria obtained with the best hANN models for each catchment during calibration period.

Catchment	hANN	MSE	NS	PI	cAI1	cAI2
01127000	6-6-AdaptPSO-MSE	1.23×10^{-1}	8.78×10^{-1}	-2.82×10^{-1}	8.97×10^{-1}	6.00×10^{-1}
	6-6-AdaptPSO-PI	1.13×10^{-1}	8.78×10^{-1}	-2.93×10^{-1}	8.97×10^{-1}	6.00×10^{-1}
	6-6-AdaptPSO-cAI1	1.13×10^{-1}	8.78×10^{-1}	-2.82×10^{-1}	9.86×10^{-1}	6.00×10^{-1}
	6-6-AdaptPSO-cAI2	1.13×10^{-1}	8.78×10^{-1}	-2.82×10^{-1}	8.97×10^{-1}	6.30×10^{-1}
01197500	12*s*-6-APartPSO-NS	9.64×10^{-2}	8.98×10^{-1}	-2.43×10^{-1}	8.47×10^{-1}	5.16×10^{-1}
01321000	6-6-AdaptPSO-NS	9.42×10^{-2}	8.93×10^{-1}	-2.25×10^{-1}	8.70×10^{-1}	5.37×10^{-1}
01371500	12*s*-6-APartPSO-NS	9.42×10^{-2}	9.09×10^{-1}	-2.55×10^{-1}	8.47×10^{-1}	4.96×10^{-1}
01372500	12*s*-6-NonlinPSO-NS	1.08×10^{-1}	8.98×10^{-1}	-3.17×10^{-1}	9.30×10^{-1}	5.38×10^{-1}
01426500	12*s*-6-APartPSO-NS	1.05×10^{-1}	8.83×10^{-1}	-2.65×10^{-1}	8.89×10^{-1}	5.43×10^{-1}
01445500	12*s*-6-APartPSO-NS	9.85×10^{-2}	9.02×10^{-1}	-1.74×10^{-1}	8.68×10^{-1}	5.13×10^{-1}
01503000	6-3-APartPSO-MSE	1.28×10^{-1}	8.67×10^{-1}	-8.04×10^{-2}	9.15×10^{-1}	3.08×10^{-1}

Categorical skill scores (Fig. A1) show that hANN performs comparably to persistence at the mild drought threshold (SPEI ≤ -1.0), but categorical skill deteriorates at rarer, more severe drought levels (SPEI $\leq -1.5, -2.0$). This indicates that while hANN improves continuous prediction metrics (NSE), persistence remains competitive when predictions are discretized into drought categories.

4. Discussion

Our results indicate that the number of neurons in the hidden layer plays a key role in model performance. The optimal configuration in this study was $N_{hd} = 6$, which provided the best results for both calibration and validation. This finding deviates from the heuristic proposed by Wanas et al. (1998), who suggest $N_{hd} = \log(T)$, where T is the number of training samples. In our case, with $T = 324$, this would result in $N_{hd} = 3$, a configuration that led to substantially worse performance.

Other common heuristics, such as $N_{hd} = 2n + 1$ proposed by Lippmann (1987) and Mishra and Desai (2006), would yield values of 13 or 25 for our models with 6 or 12 input nodes, respectively. However, we found that smaller hidden layers than those recommended by these formulas, specifically fewer neurons than input nodes, can still provide strong performance

TABLE 9. The best achieved accuracy criteria obtained with the best hANN models for each catchment during validation period.

Catchment	hANN	MSE	NS	PI	cAI1	cAI2
01127000	6-6-AdaptPSO-MSE	1.76×10^{-1}	7.58×10^{-1}	-1.78×10^{-1}	8.96×10^{-1}	7.46×10^{-1}
	6-6-AdaptPSO-PI	1.76×10^{-1}	7.58×10^{-1}	-1.18×10^{-1}	8.96×10^{-1}	7.46×10^{-1}
	6-6-AdaptPSO-cAI1	1.76×10^{-1}	7.58×10^{-1}	-1.78×10^{-1}	9.07×10^{-1}	7.46×10^{-1}
	6-6-AdaptPSO-cAI2	1.76×10^{-1}	7.58×10^{-1}	-1.78×10^{-1}	8.96×10^{-1}	7.29×10^{-1}
01197500	12-6-APartPSO-MSE	1.65×10^{-1}	7.94×10^{-1}	-1.36×10^{-1}	9.30×10^{-1}	6.83×10^{-1}
	12-6-APartPSO-PI	1.61×10^{-1}	7.94×10^{-1}	-2.31×10^{-1}	9.30×10^{-1}	6.83×10^{-1}
	12-6-APartPSO-cAI1	1.61×10^{-1}	7.94×10^{-1}	-1.36×10^{-1}	9.98×10^{-1}	6.83×10^{-1}
	12-6-APartPSO-cAI2	1.61×10^{-1}	7.94×10^{-1}	-1.36×10^{-1}	9.30×10^{-1}	7.01×10^{-1}
01321000	12*s*-6-LinPSO-NS	1.45×10^{-1}	8.07×10^{-1}	-3.08×10^{-1}	9.04×10^{-1}	6.48×10^{-1}
01371500	6-6-APartPSO-MSE	1.62×10^{-1}	7.95×10^{-1}	-1.27×10^{-1}	9.25×10^{-1}	6.99×10^{-1}
	6-6-APartPSO-PI	1.51×10^{-1}	7.95×10^{-1}	-1.36×10^{-1}	9.25×10^{-1}	6.99×10^{-1}
	6-6-APartPSO-cAI1	1.51×10^{-1}	7.95×10^{-1}	-1.27×10^{-1}	9.65×10^{-1}	6.99×10^{-1}
	6-6-APartPSO-cAI2	1.51×10^{-1}	7.95×10^{-1}	-1.27×10^{-1}	9.25×10^{-1}	6.87×10^{-1}
01372500	12*s*-6-APartPSO-MSE	1.76×10^{-1}	7.71×10^{-1}	-2.06×10^{-1}	9.15×10^{-1}	7.16×10^{-1}
	12*s*-6-APartPSO-PI	1.64×10^{-1}	7.71×10^{-1}	-1.76×10^{-1}	9.15×10^{-1}	7.16×10^{-1}
	12*s*-6-APartPSO-cAI1	1.64×10^{-1}	7.71×10^{-1}	-2.06×10^{-1}	9.74×10^{-1}	7.16×10^{-1}
	12*s*-6-APartPSO-cAI2	1.64×10^{-1}	7.71×10^{-1}	-2.06×10^{-1}	9.15×10^{-1}	7.40×10^{-1}
01426500	6-6-NonlinPSO-MSE	1.72×10^{-1}	8.23×10^{-1}	-2.81×10^{-2}	9.22×10^{-1}	6.87×10^{-1}
	6-6-NonlinPSO-PI	1.56×10^{-1}	8.23×10^{-1}	-2.14×10^{-1}	9.22×10^{-1}	6.87×10^{-1}
	6-6-NonlinPSO-cAI1	1.56×10^{-1}	8.23×10^{-1}	-2.81×10^{-2}	9.91×10^{-1}	6.87×10^{-1}
	6-6-NonlinPSO-cAI2	1.56×10^{-1}	8.23×10^{-1}	-2.81×10^{-2}	9.22×10^{-1}	7.18×10^{-1}
01445500	12*s*-6-ChaoPSO-MSE	1.38×10^{-1}	8.24×10^{-1}	-1.45×10^{-1}	9.23×10^{-1}	6.20×10^{-1}
	12*s*-6-ChaoPSO-PI	1.31×10^{-1}	8.24×10^{-1}	-2.00×10^{-1}	9.23×10^{-1}	6.20×10^{-1}
	12*s*-6-ChaoPSO-cAI1	1.31×10^{-1}	8.24×10^{-1}	-1.45×10^{-1}	9.62×10^{-1}	6.20×10^{-1}
	12*s*-6-ChaoPSO-cAI2	1.31×10^{-1}	8.24×10^{-1}	-1.45×10^{-1}	9.23×10^{-1}	6.29×10^{-1}
01503000	6-6-NonlinPSO-MSE	1.77×10^{-1}	8.26×10^{-1}	-1.52×10^{-1}	9.28×10^{-1}	7.05×10^{-1}
	6-6-NonlinPSO-PI	1.64×10^{-1}	8.26×10^{-1}	-1.53×10^{-1}	9.28×10^{-1}	7.05×10^{-1}
	6-6-NonlinPSO-cAI1	1.64×10^{-1}	8.26×10^{-1}	-1.52×10^{-1}	9.63×10^{-1}	7.05×10^{-1}
	6-6-NonlinPSO-cAI2	1.64×10^{-1}	8.26×10^{-1}	-1.52×10^{-1}	9.28×10^{-1}	6.97×10^{-1}

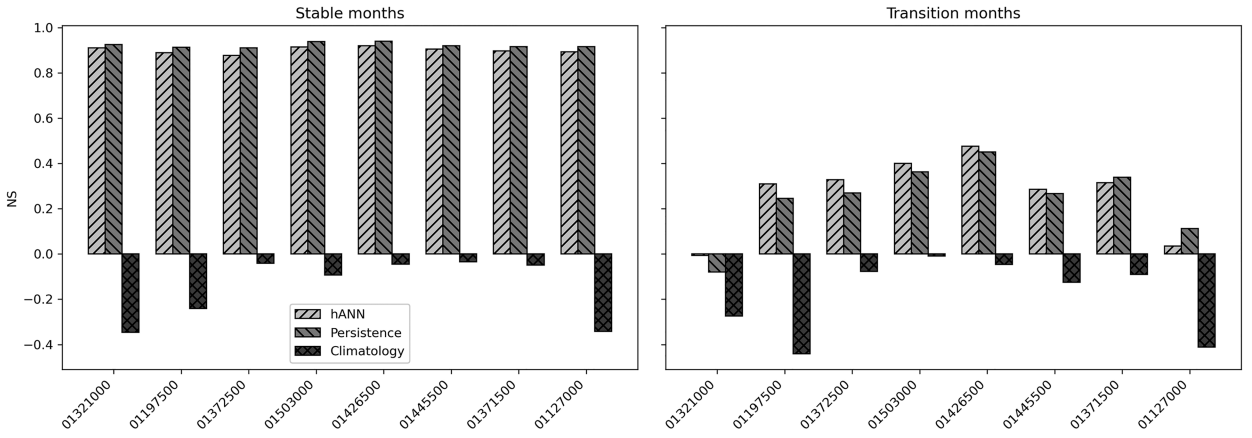

FIG. 7. The NS of hANN, persistence, and seasonal climatology models for each catchment, calculated separately for stable ($\Delta SPEI_t < 0.5$) and transition ($\Delta SPEI_t \geq 0.5$) months in the validation period.

while reducing the number of parameters to optimize, which is consistent with findings from Fletcher and Goss (1993) and Zhang and Dong (2001).

Among the five PSO variants tested, the APartPSO algorithm clearly outperformed the others. It produced significantly better results across both calibration and validation periods, indicating its superior ability to explore complex, multidimensional solution spaces. This aligns with previous research where APartPSO showed strong performance in benchmark function optimization and rainfall–runoff model calibration (Jakubcová et al. 2015). These results highlight its robustness and suitability for real-world hydrological applications.

In terms of objective functions, the NS consistently led to better model calibration and validation compared to the other tested criteria. The NS metric, along with MSE, remains a widely accepted and reliable performance indicator in hydrological and meteorological modeling. Our best-performing model achieved MSE values of 0.093 and 0.130 and NS values of 0.909 and 0.826 for calibration and validation, respectively. These results are competitive with existing ANN-based drought prediction studies. For example, Hosseini-Moghari and Araghinejad (2015) reported MSE = 0.144 and NS = 0.834 for SPI simulations, while Belayneh and Adamowski (2012) and Deo and Şahin (2015) achieved higher accuracies (e.g., NS up to 0.953 and 0.983), although their models used significantly larger hidden layers, up to 43 neurons, which likely contributed to their improved performance.

The PI will be negative when the proposed model performs worse than the naive persistence forecast in terms of NS. This typically occurred during the stable months regime, where high month-to-month autocorrelation in SPEI meant that the most recent observation was already a strong predictor of the next value. Under such conditions, even a well-trained ANN may offer little improvement over persistence, and small errors can lead to lower skill scores than the baseline.

Negative PI values were also observed for some catchments during the transition months regime. In these cases, the hANN did not consistently outperform persistence, possibly because rapid hydroclimatic shifts were not fully captured by the available predictors. From an operational perspective, this indicates that while the hANN can match or exceed persistence skill in certain situations, particularly during transitions for some catchments, it does not guarantee improvement across all regimes. Where persistence already delivers high skill, the additional complexity of an ANN must be weighed against its marginal benefit. Conversely, in applications where accurate anticipation of rapid changes or extremes is critical, targeted model refinements focused on these high-variability periods could yield meaningful gains in utility.

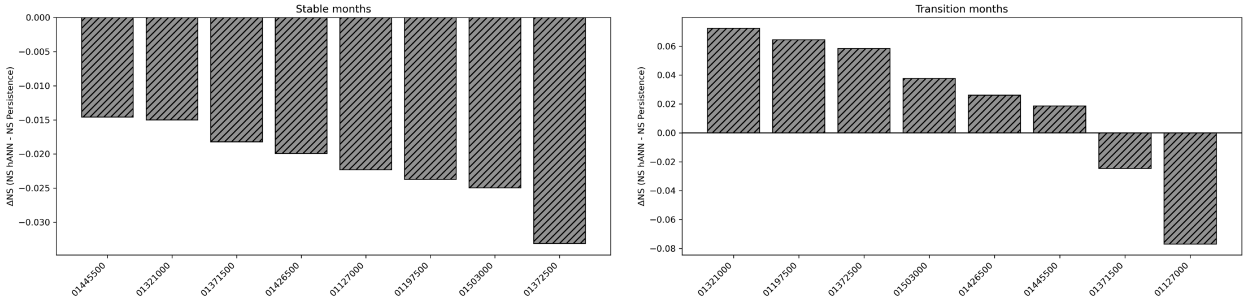

FIG. 8. Difference in NS between hANN and persistence for stable and transition months.

5. Conclusions

This study aimed to improve the forecasting of the SPEI by integrating hANN models with PSO algorithms for model weight training. A total of 150 model configurations were tested across eight U.S. catchments using 54 years of daily meteorological data. The models were evaluated during both the calibration and validation periods using five different goodness-of-fit metrics and were compared against two standard baseline approaches—persistence and seasonal climatology.

The best-performing model configuration used six input variables, six hidden neurons, the APartPSO optimization algorithm, and the NS as the objective function. This setup achieved strong performance while maintaining computational efficiency, confirming the benefit of moderate network complexity when combined with a well-adapted optimizer.

Among the tested PSO variants, APartPSO consistently delivered the most accurate and stable results. However, the study also revealed that the effectiveness of a given optimization method is not universal, performance can vary depending on the case study and model structure. This highlights the importance of testing multiple optimizers and configurations when applying hybrid models in hydrometeorological forecasting.

The analysis also showed that hANN skill gains over persistence were both catchment- and regime-dependent, with negative PI values observed in some validation cases. In particular, persistence remained a strong predictor under stable hydroclimatic conditions, limiting the potential improvement from a more complex model.

Overall, the findings highlight that the effectiveness of a given optimization method or model configuration can vary depending on both catchment characteristics and prevailing hydroclimatic regimes. While the hANN demonstrated potential to exceed baseline skill, especially during transitional periods for certain catchments, its operational benefit should be evaluated in light of baseline performance.

Future research should explore the sensitivity of hANN parameters and investigate alternative network architectures, activation functions, and objective functions. It would also be valuable to assess model's generalizability to other hydrometeorological contexts, with particular attention to performance during periods of rapid changes or extreme events.

Acknowledgments. The computational resources were provided by the e-INFRA CZ project (ID:90254), supported by the Ministry of Education, Youth and Sports of the Czech Republic. We also acknowledge support from the Operational Programme Just Transition (CZ.10.01.01/00/22_001/0000287 "Smart Landscape 2030+"). We further appreciate the institutional support of the Institute of Thermomechanics of the Czech Academy of Sciences for its long-term conceptual development (RVO: 61388998). We are grateful to the anonymous referees for their valuable comments and suggestions, which helped improve the presentation of this paper.

Data availability statement. The dataset used in this study is available at the NOAA Hydrology website (https://hydrology.nws.noaa.gov/pub/). The source code of the model is available on GitHub (https://github.com/jakubcovam/MLP-PSO). Due to their size, the raw model simulation outputs are not easily archived or shared; however, we have provided all information necessary to replicate the simulations, and these outputs can be made available upon request to the corresponding author.

APPENDIX

Categorical Drought-Skill Analysis

a. Methodology

To complement continuous skill evaluation (e.g., NSE), we evaluated categorical drought prediction skill by discretizing both observed and predicted SPEI values into drought classes. Three severity thresholds were considered, corresponding to moderate drought (SPEI ≤ -1.0), severe drought (SPEI ≤ -1.5), and extreme drought (SPEI ≤ -2.0).

For each threshold, both observed and predicted SPEI values were transformed into binary event series (1 = drought, 0 = no drought). Model performance was then quantified using contingency tables that classify outcomes into a: hits (observed and predicted droughts), b: false alarms (predicted but not observed), c: misses (observed but not predicted), and d: correct negatives (both predict and observe no drought). The total number of cases N is then defined as

$$N = a + b + c + d. \tag{A1}$$

From these counts, three categorical verification metrics were derived (Wilks 2011; Su et al. 2023).

Hit rate (HR) is

$$\mathrm{HR} = \frac{a}{a + c}, \tag{A2}$$

which expresses the fraction of observed drought events that were successfully predicted.

False alarm ratio (FAR) is

$$\mathrm{FAR} = \frac{b}{a + b}, \tag{A3}$$

which measures the fraction of predicted droughts that did not materialize.

Heidke skill score (HSS) is

$$\mathrm{HSS} = \frac{2(ad - bc)}{(a + c)(c + d) + (a + b)(b + d)}, \tag{A4}$$

which evaluates the overall categorical skill relative to random chance. A value of HSS = 1 indicates perfect forecasts, HSS = 0 corresponds to no skill, and negative values indicate performance worse than chance.

This categorical evaluation complements the continuous skill assessment by providing insight into the models' ability to capture drought occurrence in a practically relevant classification framework. Since the number of drought events decreases with increasing severity, the robustness of categorical metrics diminishes at higher thresholds, making results for SPEI ≤ -2.0 particularly sensitive to sample size limitations.

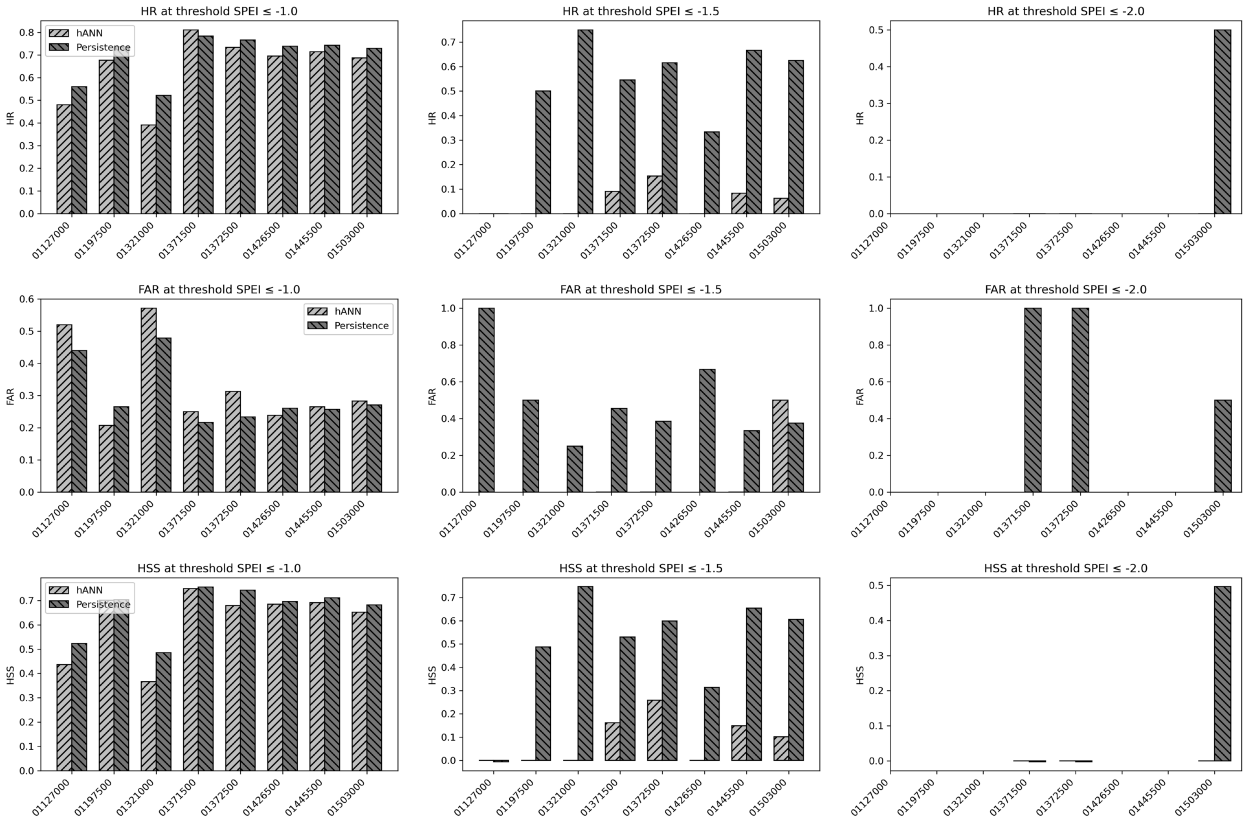

FIG. A1. Categorical drought prediction skill of the hANN model compared with the persistence baseline for all eight catchments. Performance is shown for three drought thresholds based on SPEI: (left) moderate drought, (center) severe drought, and (right) extreme drought. Rows correspond to different verification metrics: (top) HR, (middle) FAR, and (bottom) HSS.

In addition to categorical metrics, model performance was evaluated using receiver operating characteristic (ROC) analysis, which provides a threshold-independent view of classification skill. ROC analysis quantifies the trade-off between correctly detecting drought events and avoiding false alarms across a continuous range of decision thresholds.

For a given probability threshold, model predictions are binarized and classified into hits, false alarms, misses, and correct negatives as defined above. From these counts, two rates are derived. The true positive rate (TPR) also known as hit rate or sensitivity is in Eq. (A2), and the false positive rate (FPR) is

$$\mathrm{FPR} = \frac{b}{b + d}, \tag{A5}$$

which denotes the fraction of nondrought cases incorrectly classified as droughts.

The ROC curve is obtained by plotting TPR against FPR for all possible discrimination thresholds. A model with no skill produces a curve along the 1:1 diagonal (i.e., TPR = FPR), while a perfect model follows the top-left border of the ROC space (TPR = 1 and FPR = 0).

The area under the ROC curve (AUC) provides a scalar summary of the ROC curve as

$$\mathrm{AUC} = \int_0^1 \mathrm{TPR}(\mathrm{FPR})d(\mathrm{FPR}). \tag{A6}$$

AUC values range from 0.5 (no discrimination skill, equivalent to random guessing) to 1.0 (perfect discrimination). Intermediate values quantify the probability that the model will rank a randomly chosen drought event higher than a nonevent.

b. Results and conclusions

The categorical drought-skill analysis is displayed in Fig. A1. At the moderate drought threshold (left column), hANN achieved similar or slightly better categorical skill compared to persistence across most catchments. HR values were generally between 0.5 and 0.8, while FAR ranged from 0.2 to 0.5. HSS values confirmed that both hANN and persistence provided useful categorical information.

At the severe threshold (middle column), categorical skill declined substantially for hANN, with HR typically <0.2 and HSS close to zero. Persistence remained competitive, reflecting its strong ability to carry information across consecutive months in rare-event settings.

At the extreme threshold (right column), the scarcity of events led to very sparse statistics, with categorical scores dominated by persistence and hANN often failing to register hits. These results suggest limited reliability of categorical predictions for very rare droughts, independent of the modeling approach.

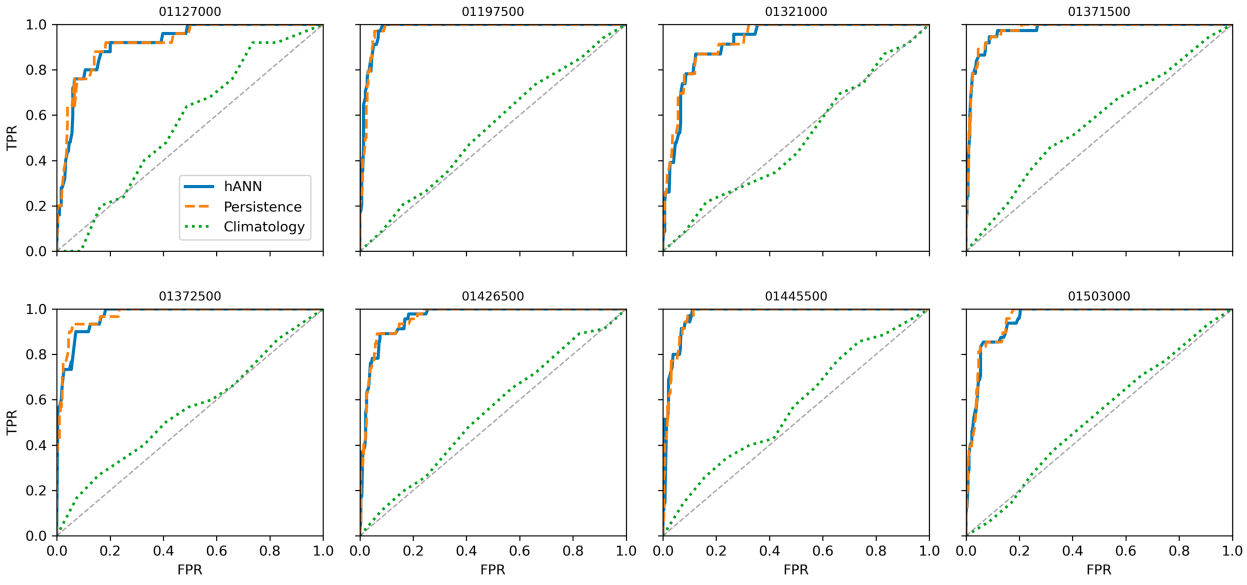

FIG. A2. ROC curves for drought detection (SPEI ≤ −1.0) across all eight catchments. The solid blue, dashed orange, and dotted green curves denote the hANN, the persistence, and the seasonal climatology, respectively. The dashed diagonal line indicates the no-skill reference (AUC = 0.5).

The seasonal climatology baseline is not shown in the categorical plots because it consistently produced HR and HSS values close to zero, corresponding to no skill beyond random chance. This outcome reflects the inability of a purely seasonal predictor to capture individual drought events, particularly under the more stringent thresholds, and justifies focusing the comparison on hANN and persistence.

To complement the categorical verification metrics, Figs. A2–A4 show the ROC curves for drought events. Across all catchments, both the hANN and the persistence benchmark exhibit ROC curves that rise well above the no-skill line, confirming their ability to discriminate drought from non-drought states. The AUC is consistently high for both methods, typically exceeding 0.8, indicating strong discrimination skill. The hANN curves are generally comparable to persistence, with slightly higher TPRs at moderate FPRs in several catchments, suggesting modest added value of the hybrid learning approach over the benchmark.

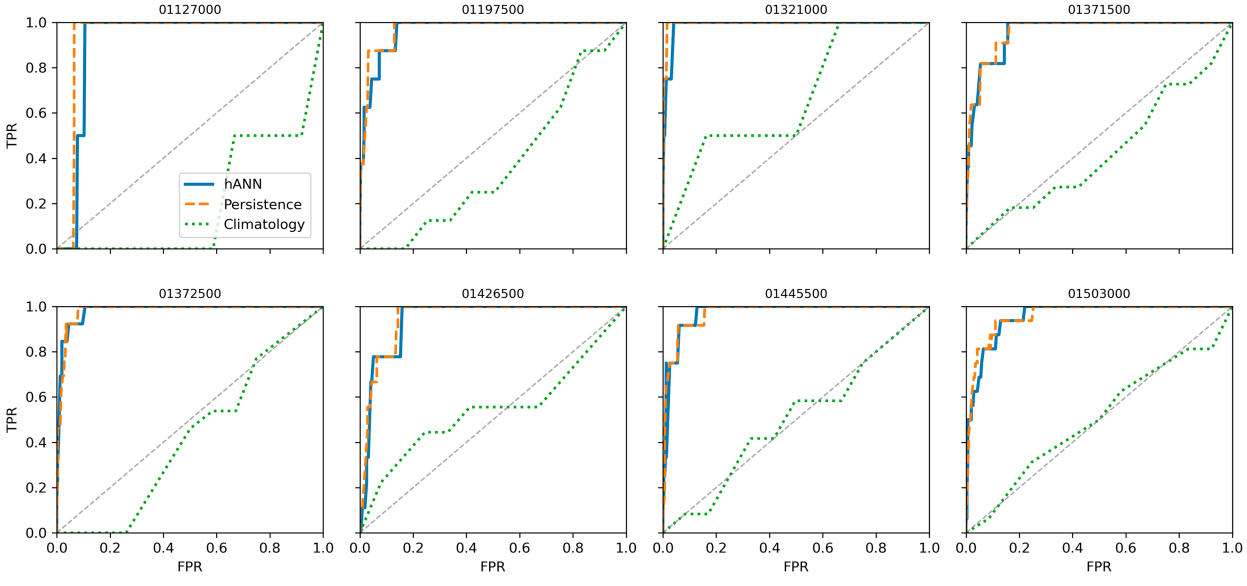

FIG. A3. ROC curves for drought detection (SPEI ≤ −1.5) across all eight catchments. The solid blue, dashed orange, and dotted green curves denote the hANN, the persistence, and the seasonal climatology, respectively. The dashed diagonal line indicates the no-skill reference (AUC = 0.5).

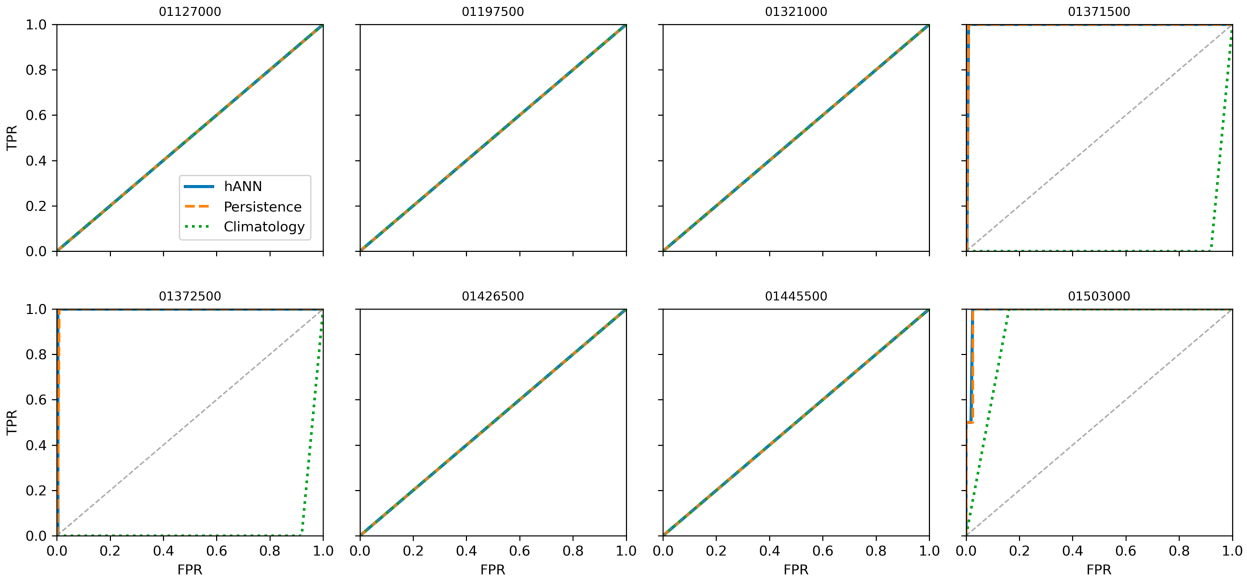

FIG. A4. ROC curves for drought detection (SPEI ≤ −2.0) across all eight catchments. The solid blue, dashed orange, and dotted green curves denote the hANN, the persistence, and the seasonal climatology, respectively. The dashed diagonal line indicates the no-skill reference (AUC = 0.5).

In contrast, the climatology baseline exhibited much weaker performance, with ROC curves lying only slightly above the diagonal, reflecting its limited ability to capture interannual variability in drought occurrence. This finding is consistent with the categorical skill metrics (HR and HSS close to zero) observed across all thresholds (not shown).

The categorical analysis highlights an important distinction between continuous predictive skill (e.g., NSE) and drought-classification performance. While hANN clearly improved continuous prediction over climatology and persistence, categorical skill did not translate as strongly, especially for rare and extreme droughts. This discrepancy is likely due to the data imbalance inherent in drought time series.

The ROC analysis corroborates the categorical drought classification results. The hANN achieves skill comparable to persistence, with improvements in several catchments, while climatology remains uninformative for drought onset. Importantly, the ROC framework highlights that these conclusions are robust across a range of decision thresholds, not only at the conventional categorical cutoffs.

These findings emphasize that persistence remains a strong baseline in categorical drought detection, particularly at higher severity levels. Future work could explore imbalance-aware training strategies (e.g., weighted loss functions) to improve categorical drought skill.

REFERENCES

Al-Juboori, A. M., 2023: Prediction of hydrological drought in semi-arid regions using a novel hybrid model. *Water Resour. Manage.*, **37**, 3657–3669, https://doi.org/10.1007/s11269-023-03520-1.

Beguería, S., and S. M. Vicente-Serrano, 2013: SPEI: Calculation of the standardised precipitation-evapotranspiration index, version 1.6. R Package, https://github.com/sbegueria/SPEI.

——, ——, F. Reig, and B. Latorre, 2014: Standardized precipitation evapotranspiration index (SPEI) revisited: Parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. *Int. J. Climatol.*, **34**, 3001–3023, https://doi.org/10.1002/joc.3887.

Belayneh, A., and J. Adamowski, 2012: Standard precipitation index drought forecasting using neural networks, wavelet neural networks, and support vector regression. *Appl. Comput. Intell. Soft Comput.*, **2012**, 794061, https://doi.org/10.1155/2012/794061.

Bennett, N. D., and Coauthors, 2013: Characterising performance of environmental models. *Environ. Modell. Software*, **40**, 1–20, https://doi.org/10.1016/j.envsoft.2012.09.011.

Chau, K. W., 2006: Particle swarm optimization training algorithm for ANNs in stage prediction of Shing Mun River. *J. Hydrol.*, **329**, 363–367, https://doi.org/10.1016/j.jhydrol.2006.02.025.

——, 2007: A split-step particle swarm optimization algorithm in river stage forecasting. *J. Hydrol.*, **346**, 131–135, https://doi.org/10.1016/j.jhydrol.2007.09.004.

da S. Gomes, G. S., T. B. Ludermir, and L. M. M. R. Lima, 2011: Comparison of new activation functions in neural network for forecasting financial time series. *Neural Comput. Appl.*, **20**, 417–439, https://doi.org/10.1007/s00521-010-0407-3.

Dawson, C. W., R. J. Abrahart, and L. M. See, 2007: HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Environ. Modell. Software*, **22**, 1034–1052, https://doi.org/10.1016/j.envsoft.2006.06.008.

Deo, R. C., and M. Şahin, 2015: Application of the artificial neural network model for prediction of monthly standardized precipitation and evapotranspiration index using hydrometeorological parameters and climate indices in eastern Australia.

Atmos. Res., **161–162**, 65–81, https://doi.org/10.1016/j.atmosres.2015.03.018.

Dikshit, A., B. Pradhan, and A. M. Alamri, 2021: Long lead time drought forecasting using lagged climate variables and a stacked long short-term memory model. *Sci. Total Environ.*, **755**, 142638, https://doi.org/10.1016/j.scitotenv.2020.142638.

Duan, Q., and Coauthors, 2006: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *J. Hydrol.*, **320**, 3–17, https://doi.org/10.1016/j.jhydrol.2005.07.031.

Duch, W., and N. Jankowski, 1999: Survey of neural transfer functions. *Neural Comput. Surv.*, **2**, 163–212.

Eberhart, R. C., and Y. Shi, 2000: Comparing inertia weights and constriction factors in particle swarm optimization. *Proc. 2000 Congress on Evolutionary Computation*, La Jolla, CA, Institute of Electrical and Electronics Engineers, 84–88, https://doi.org/10.1109/CEC.2000.870279.

——, and ——, 2007: *Computational Intelligence: Concepts to Implementations.* Elsevier, 467 pp.

——, P. K. Simpson, and R. C. Dobbins, 1996: *Computational Intelligence PC Tools.* Academic Press Professional, Inc., 464 pp.

Ehteram, M., M. Achite, Z. Sheikh Khozani, and A. Farrokhi, 2024: Drought prediction using ensemble models. *Acta Geophys.*, **72**, 945–982, https://doi.org/10.1007/s11600-023-01058-9.

Elbeltagi, A., and Coauthors, 2024: Advanced stacked integration method for forecasting long-term drought severity: CNN with machine learning models. *J. Hydrol.*, **53**, 101759, https://doi.org/10.1016/j.ejrh.2024.101759.

Esquivel-Saenz, P. J., R. Ortiz-Gómez, M. Zavala, and R. S. Flowers-Cano, 2024: Artificial neural networks for drought forecasting in the central region of the state of Zacatecas, Mexico. *Climate*, **12**, 131, https://doi.org/10.3390/cli12090131.

Feng, Y., G.-F. Teng, A.-X. Wang, and Y.-M. Yao, 2007: Chaotic inertia weight in particle swarm optimization. *Second Int. Conf. on Innovative Computing, Information and Control (ICICIC 2007)*, Kumamoto, Japan, Institute of Electrical and Electronics Engineers, 475, https://doi.org/10.1109/ICICIC.2007.209.

Fletcher, D., and E. Goss, 1993: Forecasting with neural networks: An application using bankruptcy data. *Inf. Manage.*, **24**, 159–167, https://doi.org/10.1016/0378-7206(93)90064-Z.

Fu, Z., and Coauthors, 2024: Global critical soil moisture thresholds of plant water stress. *Nat. Commun.*, **15**, 4826, https://doi.org/10.1038/s41467-024-49244-7.

Goswami, M., K. M. O'Connor, K. P. Bhattarai, and A. Y. Shamseldin, 2005: Assessing the performance of eight real-time updating models and procedures for the Brosna River. *Hydrol. Earth Syst. Sci.*, **9**, 394–411, https://doi.org/10.5194/hess-9-394-2005.

Granata, F., F. Di Nunno, M. Najafzadeh, and I. Demir, 2022: A stacked machine learning algorithm for multi-step ahead prediction of soil moisture. *Hydrology*, **10** (1), 1, https://doi.org/10.3390/hydrology10010001.

Hanel, M., O. Rakovec, Y. Markonis, P. Maca, L. Samaniego, J. Kysely, and R. Kumar, 2018: Revisiting the recent European droughts from a long-term perspective. *Sci. Rep.*, **8**, 9499, https://doi.org/10.1038/s41598-018-27464-4.

Haslinger, K., D. Koffler, W. Schöner, and G. Laaha, 2014: Exploring the link between meteorological drought and streamflow: Effects of climate-catchment interaction. *Water Resour. Res.*, **50**, 2468–2487, https://doi.org/10.1002/2013WR015051.

Hayes, M., M. Svoboda, N. Wall, and M. Widhalm, 2011: The Lincoln declaration on drought indices: Universal meteorological drought index recommended. *Bull. Amer. Meteor. Soc.*, **92**, 485–488, https://doi.org/10.1175/2010BAMS3103.1.

He, X., H. Guan, and J. Qin, 2015: A hybrid wavelet neural network model with mutual information and particle swarm optimization for forecasting monthly rainfall. *J. Hydrol.*, **527**, 88–100, https://doi.org/10.1016/j.jhydrol.2015.04.047.

Hernandez, E. A., and V. Uddameri, 2014: Standardized precipitation evaporation index (SPEI)-based drought assessment in semi-arid South Texas. *Environ. Earth Sci.*, **71**, 2491–2501, https://doi.org/10.1007/s12665-013-2897-7.

Hornik, K., M. Stinchcombe, and H. White, 1989: Multilayer feedforward networks are universal approximators. *Neural Networks*, **2**, 359–366, https://doi.org/10.1016/0893-6080(89)90020-8.

Hosseini-Moghari, S., and S. Araghinejad, 2015: Monthly and seasonal drought forecasting using statistical neural networks. *Environ. Earth Sci.*, **74**, 397–412, https://doi.org/10.1007/s12665-015-4047-x.

Huo, Z., S. Feng, S. Kang, G. Huang, F. Wang, and P. Guo, 2012: Integrated neural networks for monthly river flow estimation in arid inland basin of northwest China. *J. Hydrol.*, **420–421**, 159–170, https://doi.org/10.1016/j.jhydrol.2011.11.054.

Jakubcová, M., P. Máca, and P. Pech, 2014: A comparison of selected modifications of the particle swarm optimization algorithm. *J. Appl. Math.*, **2014**, 293087, https://doi.org/10.1155/2014/293087.

——, ——, and ——, 2015: Parameter estimation in rainfall-runoff modelling using distributed versions of particle swarm optimization algorithm. *Math. Probl. Eng.*, **2015**, 968067, https://doi.org/10.1155/2015/968067.

Karbasi, M., M. Karbasi, M. Jamei, A. Malik, and H. M. Azamathulla, 2022: Development of a new wavelet-based hybrid model to forecast multi-scalar SPEI drought index (case study: Zanjan City, Iran). *Theor. Appl. Climatol.*, **147**, 499–522, https://doi.org/10.1007/s00704-021-03825-4.

Kennedy, J., and R. Eberhart, 1995: Particle swarm optimization. *Proc. ICNN'95—Int. Conf. on Neural Networks*, Perth, Western Australia, Australia, Institute of Electrical and Electronics Engineers, 1942–1948, https://doi.org/10.1109/ICNN.1995.488968.

——, R. C. Eberhart, and Y. Shi, 2001: *Swarm Intelligence.* Elsevier, 512 pp.

Kisi, O., A. Keshavarzi, J. Shiri, M. Zounemat-Kermani, and E.-S. E. Omran, 2017: Groundwater quality modeling using neuro-particle swarm optimization and neuro-differential evolution techniques. *Hydrol. Res.*, **48**, 1508–1519, https://doi.org/10.2166/nh.2017.206.

Lippmann, R., 1987: An introduction to computing with neural nets. *IEEE ASSP Mag.*, **4**, 4–22, https://doi.org/10.1109/MASSP.1987.1165576.

Máca, P., and P. Pech, 2015: The inertia weight updating strategies in particle swarm optimisation based on the beta distribution. *Math. Probl. Eng.*, **2015**, 790465, https://doi.org/10.1155/2015/790465.

——, and ——, 2016: Forecasting SPEI and SPI drought indices using the integrated artificial neural networks. *Comput. Intell. Neurosci.*, **2016**, 3868519, https://doi.org/10.1155/2016/3868519.

——, ——, and J. Pavlásek, 2014: Comparing the selected transfer functions and local optimization methods for neural network flood runoff forecast. *Math. Probl. Eng.*, **2014**, 782351, https://doi.org/10.1155/2014/782351.

May, R. M., 1976: Simple mathematical models with very complicated dynamics. *Nature*, **261**, 459–467, https://doi.org/10.1038/261459a0.

McKee, T. B., N. J. Doesken, J. Kleist, and Coauthors, 1993: The relationship of drought frequency and duration to time scales. *Proc. Eighth Conf. on Applied Climatology*, Boston, MA, Amer. Meteor. Soc., 179–183, https://climate.colostate.edu/pdfs/relationshipofdroughtfrequency.pdf.

Mishra, A. K., and V. R. Desai, 2006: Drought forecasting using feed-forward recursive neural network. *Ecol. Modell.*, **198**, 127–138, https://doi.org/10.1016/j.ecolmodel.2006.04.017.

Mohammadpour, R., Z. Asaie, M. R. Shojaeian, and M. Sadeghzadeh, 2018: A hybrid of ANN and CLA to predict rainfall. *Arabian J. Geosci.*, **11**, 533, https://doi.org/10.1007/s12517-018-3804-z.

Mossad, A., and A. A. Alazba, 2015: Drought forecasting using stochastic models in a hyper-arid climate. *Atmosphere*, **6**, 410–430, https://doi.org/10.3390/atmos6040410.

Mpelasoka, F., K. Hennessy, R. Jones, and B. Bates, 2008: Comparison of suitable drought indices for climate change impacts assessment over Australia towards resource management. *Int. J. Climatol.*, **28**, 1283–1292, https://doi.org/10.1002/joc.1649.

Namias, J., 1952: The annual course of month-to-month persistence in climatic anomalies. *Bull. Amer. Meteor. Soc.*, **33**, 279–285, https://doi.org/10.1175/1520-0477-33.7.279.

Nickabadi, A., M. M. Ebadzadeh, and R. Safabakhsh, 2011: A novel particle swarm optimization algorithm with adaptive inertia weight. *Appl. Soft Comput.*, **11**, 3658–3670, https://doi.org/10.1016/j.asoc.2011.01.037.

Nourani, V., and O. Kalantari, 2010: Integrated artificial neural network for spatiotemporal modeling of rainfall–runoff–sediment processes. *Environ. Eng. Sci.*, **27**, 411–422, https://doi.org/10.1089/ees.2009.0353.

Ott, R., and M. Longnecker, 2008: *An Introduction to Statistical Methods and Data Analysis.* Cengage Learning, 1273 pp.

Palmer, W. C., 1965: Meteorological drought. U.S. Weather Bureau Research Paper 45, 58 pp.

Pereira, L. S., I. Cordery, and I. Iacovides, 2009: *Coping with Water Scarcity: Addressing the Challenges.* Springer, 382 pp.

Poli, R., J. Kennedy, and T. Blackwell, 2007: Particle swarm optimization: An overview. *Swarm Intell.*, **1**, 33–57, https://doi.org/10.1007/s11721-007-0002-0.

Poudel, B., D. Dahal, M. Banjara, and A. Kalra, 2024: Assessing meteorological drought patterns and forecasting accuracy with SPI and SPEI using machine learning models. *Forecasting*, **6**, 1026–1044, https://doi.org/10.3390/forecast6040051.

R C Team, 2021: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org.

Riebsame, W. E., S. A. Changnon Jr., and T. R. Karl, 1991: *Drought and Natural Resources Management in the United States: Impacts and Implications of the 1987–89 Drought.* Westview Press Inc., 192 pp.

Rumelhart, D. E., and Coauthors, 1988: *Parallel Distributed Processing.* Vol. 1. Institute of Electrical and Electronics Engineers, 565 pp.

Shi, Y, and R. Eberhart, 1999: Empirical study of particle swarm optimization. *Proc. 1999 Congress on Evolutionary Computation-CEC99*, Washington, DC, Institute of Electrical and Electronics Engineers, 1945–1950, https://doi.org/10.1109/CEC.1999.785511.

Simon, D., 2013: *Evolutionary Optimization Algorithms.* John Wiley and Sons, 784 pp.

Singh, U., and Coauthors, 2023: Hybrid multi-model ensemble learning for reconstructing gridded runoff of Europe for 500 years. *Inf. Fusion*, **97**, 101807, https://doi.org/10.1016/j.inffus.2023.101807.

Stagge, J. H., L. M. Tallaksen, L. Gudmundsson, A. F. Van Loon, and K. Stahl, 2015: Candidate distributions for climatological drought indices (SPI and SPEI). *Int. J. Climatol.*, **35**, 4027–4040, https://doi.org/10.1002/joc.4267.

Su, L., Q. Cao, S. Shukla, M. Pan, and D. P. Lettenmaier, 2023: Evaluation of subseasonal drought forecast skill over the coastal western United States. *J. Hydrometeor.*, **24**, 709–726, https://doi.org/10.1175/JHM-D-22-0103.1.

Sun, P., C. Ge, R. Yao, Y. Bian, H. Yang, Q. Zhang, C.-Y. Xu, and V. P. Singh, 2024: Development of a Nonstationary Standardized Precipitation Evapotranspiration Index (NSPEI) and its application across China. *Atmos. Res.*, **300**, 107256, https://doi.org/10.1016/j.atmosres.2024.107256.

Tuel, A., and O. Martius, 2023: Weather persistence on subseasonal to seasonal timescales: A methodological review. *Earth Syst. Dyn.*, **14**, 955–987, https://doi.org/10.5194/esd-14-955-2023.

Vadez, V., A. Grondin, K. Chenu, A. Henry, L. Laplaze, E. J. Millet, and A. Carminati, 2024: Crop traits and production under drought. *Nat. Rev. Earth Environ.*, **5**, 211–225, https://doi.org/10.1038/s43017-023-00514-w.

Vicente-Serrano, S. M., S. Beguería, and J. I. López-Moreno, 2010: A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *J. Climate*, **23**, 1696–1718, https://doi.org/10.1175/2009JCLI2909.1.

Wable, P. S., M. K. Jha, S. Adamala, M. K. Tiwari, and S. Biswal, 2023: Application of hybrid ANN techniques for drought forecasting in the semi-arid region of India. *Environ. Monit. Assess.*, **195**, 1090, https://doi.org/10.1007/s10661-023-11631-w.

Wanas, N., G. Auda, M. Kamel, and F. Karray, 1998: On the optimal number of hidden nodes in a neural network. *Proc. IEEE Canadian Conf. on Electrical and Computer Engineering*, Waterloo, Ontario, Canada, Institute of Electrical and Electronics Engineers, 918–921, https://doi.org/10.1109/CCECE.1998.685648.

Wilhite, D. A., and M. H. Glantz, 1985: Understanding: The drought phenomenon: The role of definitions. *Water Int.*, **10**, 111–120, https://doi.org/10.1080/02508068508686328.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences.* Vol. 100. Academic Press, 704 pp.

Wu, J., J. Long, and M. Liu, 2015: Evolving RBF neural networks for rainfall prediction using hybrid particle swarm optimization and genetic algorithm. *Neurocomputing*, **148**, 136–142, https://doi.org/10.1016/j.neucom.2012.10.043.

Yan, J., H. Tiesong, H. Chongchao, W. Xianing, and G. Faling, 2007: A shuffled complex evolution of particle swarm optimization algorithm. *Adaptive and Natural Computing Algorithms*, Springer, 341–349.

Yang, Y., and Coauthors, 2024: Climate change exacerbates the environmental impacts of agriculture. *Science*, **385**, eadn3747, https://doi.org/10.1126/science.adn3747.

Zambrano-Bigiarini, M., and R. Rojas, 2013: A model-independent particle swarm optimisation software for model calibration. *Environ. Modell. Software*, **43**, 5–25, https://doi.org/10.1016/j.envsoft.2013.01.004.

Zhang, B.-L., and Z.-Y. Dong, 2001: An adaptive neural-wavelet model for short term load forecasting. *Electr. Power Syst. Res.*, **59**, 121–129, https://doi.org/10.1016/S0378-7796(01)00138-9.

Zhaoqiang, Z., and Coauthors, 2024: Recent development on drought propagation: A comprehensive review. *J. Hydrol.*, **645**, 132196, https://doi.org/10.1016/j.jhydrol.2024.132196.