

Physics-Informed Hybrid Machine Learning for One-Month-Ahead Forecasting of Meteorological Drought (SPI) in the Lake Urmia Basin, Iran

samira salari ^{1*}✉

¹ Sistan and Baluchestan Agricultural and Natural Resources Engineering Organization, Iran

*Corresponding Authors: Email: (ssalari@ymail.com)

Received: 09 January 2026

Accepted: 07 February 2026

Available online: 06 March 2026

Abstract: Meteorological drought is a recurrent hazard in north-western Iran and a major driver of the water crisis of the Lake Urmia basin. This study forecasts the 3- and 6-month Standardized Precipitation Index (SPI-3 and SPI-6) one month ahead using monthly basin-average precipitation from the Iran Meteorological Organization (2003–2023) together with temperature, pressure, wind speed and GLDAS soil moisture. A climatology-completed SPI proxy, obtained by adding the long-term mean precipitation of the target month to the known part of the accumulation window, was introduced as a physically informed baseline, and the learning models were trained to correct its residual. Multilayer perceptron (MLP), extreme gradient boosting (XGBoost), a gated recurrent unit network (GRU) and three hybrid models (MLP tuned by particle swarm optimization, MLP-PSO; XGBoost tuned by the grey wolf optimizer, XGB-GWO, and by Harris hawks optimization, XGB-HHO) were compared with persistence. The gamma distribution was fitted on the training period only to avoid information leakage. In the 2020–2023 test period the proxy alone reduced the RMSE of persistence from 0.70 to 0.46 (SPI-3) and from 0.51 to 0.36 (SPI-6). The best learning models were XGB-HHO for SPI-3 (RMSE = 0.449, NSE = 0.66) and GRU for SPI-6 (RMSE = 0.354, NSE = 0.77), which improved only marginally on the proxy. Metaheuristic tuning strongly reduced the overfitting of the untuned learners (e.g. the SPI-3 RMSE of XGBoost decreased from 0.58 to 0.45), and MLP-PSO gave the highest drought-class accuracy for SPI-3 (81%). The results show that embedding the structure of the index in the model is more important than the choice of learner for short-lead SPI forecasting.

Keywords: Drought forecasting, Standardized Precipitation Index, Lake Urmia, GRU, Particle swarm optimization, Harris hawks optimization

1. Introduction

Drought is a slowly developing natural hazard that affects water supply, agriculture, ecosystems and the economy more widely than most other disasters (Mishra and Singh, 2010). Iran, which lies mainly in arid and semi-arid climates, has experienced frequent and prolonged droughts during recent decades, and these droughts have intensified the pressure on surface water and groundwater resources (Raziei et al., 2009). In the Lake Urmia basin, the combination of dry years and increasing water use has been identified as one of the main causes of the shrinking of the lake (AghaKouchak et al., 2015). Reliable drought early warning at monthly lead times can therefore support reservoir operation, irrigation planning and the environmental water allocations that aim to restore the lake.

The Standardized Precipitation Index (SPI; McKee et al., 1993) is the drought index recommended by the World Meteorological Organization for monitoring meteorological drought (WMO, 2012). It transforms precipitation accumulated over a chosen window (e.g. 3, 6 or 12 months) into a standard normal variable, so that drought conditions can be compared across regions and time scales. Because the SPI of the next month shares most of its accumulation window with the current SPI, short-lead SPI forecasting is partly a deterministic problem and partly a problem of forecasting next

month's precipitation. This structure is often overlooked in data-driven drought studies.

Artificial neural networks were among the first data-driven tools used for SPI forecasting (Mishra and Desai, 2006), followed by wavelet-coupled neural networks and support vector regression (Belayneh et al., 2014), tree ensembles and Gaussian process regression (Elbeltagi et al., 2023). Hybrid models that couple learners with nature-inspired optimization algorithms have become particularly popular. Examples include ANN hybrids with PSO, the salp swarm and grasshopper algorithms for hydrological drought (Nabipour et al., 2020), models trained by the nomadic people optimizer for SPI forecasting in Iran (Mohamadi et al., 2020), SVR combined with the response surface method for several drought indices at Iranian stations (Piri et al., 2023), and decomposition-based hybrids (Çitakoğlu and Coşkun, 2022). Recent work also used deep feed-forward networks tuned by the war strategy optimizer for SPEI forecasting in Iran (Moghaddasi et al., 2025). These studies generally report substantial gains of hybrid over standalone models.

Two methodological issues, however, limit the interpretation of many results. First, SPI values are often calculated from the whole record before the data are split, so that the distribution parameters contain information from the test period. Second, learning

models are rarely compared with a benchmark that uses the known part of the accumulation window, which can make a naive forecast look weak and a learned model look strong. The present study addresses both issues for the Lake Urmia basin. Its objectives were (i) to characterize meteorological drought in the basin during 2003–2023, (ii) to forecast SPI-3 and SPI-6 one month ahead with MLP, XGBoost, GRU and three metaheuristic hybrids (MLP-PSO, XGB-GWO and XGB-HHO) trained as residual corrections of a climatology-completed SPI proxy, and (iii) to evaluate the models against persistence and the proxy, both as continuous forecasts and as drought-class warnings.

2. Materials and Methods

2.1. Study Area and Data

The Lake Urmia basin (about 51,760 km²) is located in north-western Iran between 44°13'–47°53' E and 35°40'–38°30' N (Fig. 1). The climate is semi-arid with a strong seasonal cycle: most precipitation falls from October to May, with a maximum in April, whereas the summer months are almost dry (Fig. 2). Monthly basin-average precipitation (P), mean air temperature (T), sea-level pressure (SLP) and wind speed (WS) from the Iran Meteorological Organization (IRIMO), and GLDAS volumetric soil moisture (SM; Rodell et al., 2004), were obtained for January 2003 – December 2023 from the public repository of Habibi and Tasouji Hassanpour (2025). The record contains 252 months without missing values.

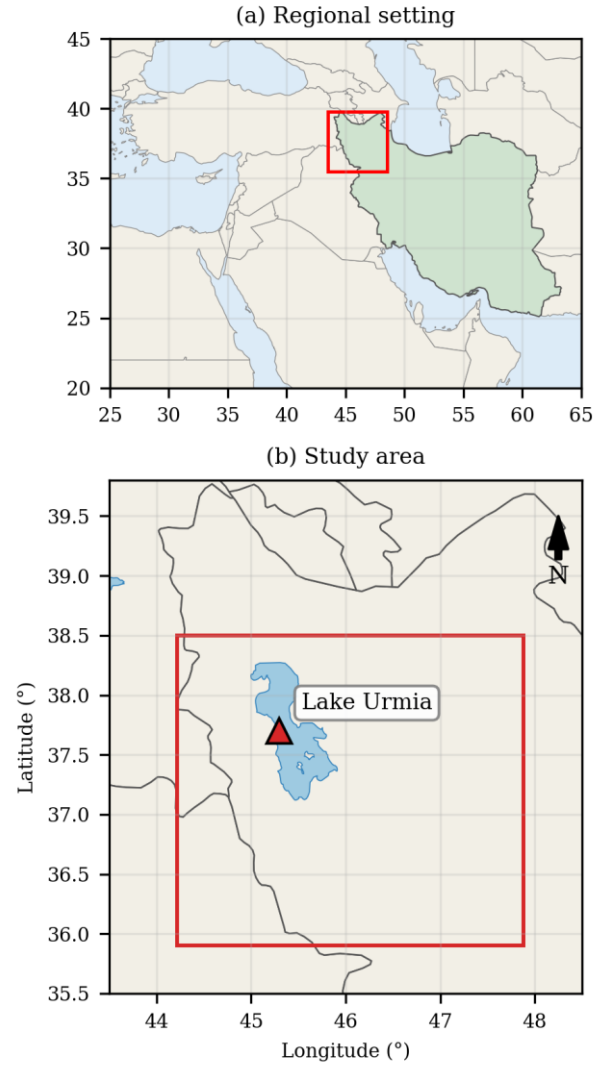


Figure 1. Location of the Lake Urmia basin (rectangle: approximate bounding box of the basin)

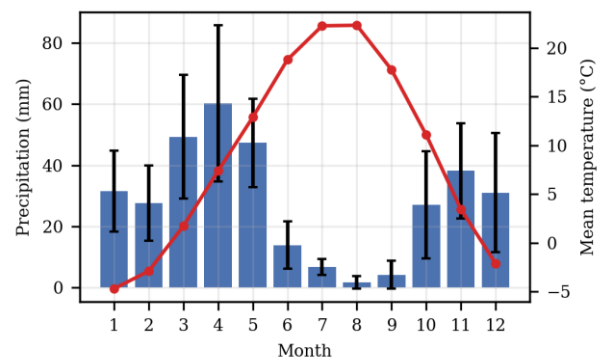


Figure 2. Mean monthly precipitation (bars, ± 1 SD) and air temperature (line) in the Lake Urmia basin, 2003–2023

Annual basin precipitation ranged from 258 mm (2008) to 506 mm (2018), with a mean of 339 mm. The Mann–Kendall test did not show a significant trend in annual precipitation ($Z = -0.92$, $p = 0.36$). Descriptive statistics of the variables are given in Table 1.

Table 1. Descriptive statistics of monthly variables and SPI series (2003–2023)

Variable	Min	Max	Mean	SD	Skew	Kurt
P (mm)	0.21	124.51	28.23	23.11	1.03	1.12
T (°C)	-9.59	24.77	9.01	9.66	-0.02	-1.35
SM (%)	10.55	42.15	23.72	7.29	0.13	-0.80
SPI-3	-2.43	3.41	-0.03	1.08	0.40	-0.30
SPI-6	-3.04	3.67	-0.00	1.12	0.40	-0.11
SPI-12	-2.60	4.75	0.05	1.34	0.63	0.77

2.2. Standardized Precipitation Index

For each accumulation window k (3, 6 and 12 months), the accumulated precipitation x of every calendar month was fitted with a two-parameter gamma distribution (McKee et al., 1993; Stagge et al., 2015):

$$g(x) = x^{\alpha-1} e^{-x/\beta} / (\beta^\alpha \Gamma(\alpha)), \quad x > 0 \quad (1)$$

where α and β are the shape and scale parameters estimated by maximum likelihood. Allowing for a probability q of zero precipitation, the cumulative probability is $H(x) = q + (1 - q)G(x)$, and the SPI is its equi-probable standard normal value:

$$SPI = \Phi^{-1}[H(x)] \quad (2)$$

where Φ is the standard normal distribution function. To prevent information leakage, α , β and q were estimated from the training period (2003–2016) only and then applied to the whole record. Drought classes followed the usual scheme: $SPI \leq -2$ extremely dry, -2 to -1.5 severely dry, -1.5 to -1 moderately dry, -1 to 1 near normal, and the symmetric wet classes (WMO, 2012). The 21-year record is shorter than the 30 years recommended for SPI calculation, which should be kept in mind when interpreting absolute SPI values.

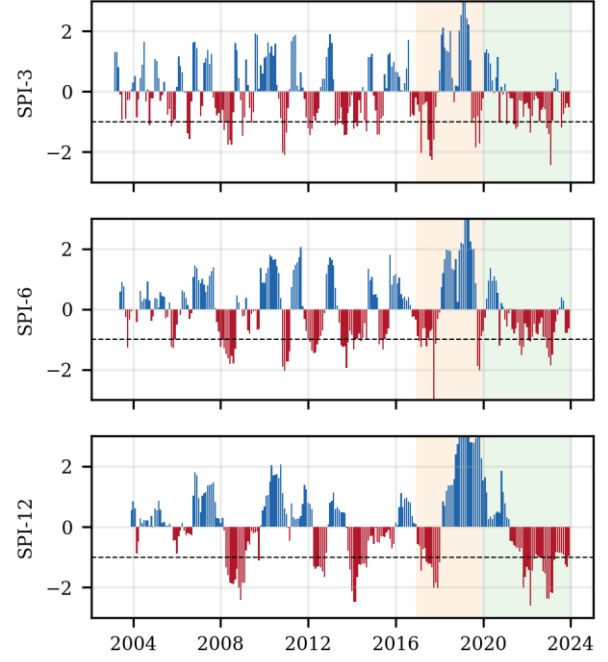


Figure 3. SPI-3, SPI-6 and SPI-12 in the Lake Urmia basin (shaded: validation 2017–2019 and test 2020–2023 periods; dashed line: SPI = -1)

2.3. Climatology-Completed SPI Proxy

At the issue month t , the precipitation of the first $k - 1$ months of the window of $SPI_k(t+1)$ is already known. A physically informed first guess is obtained by replacing the unknown precipitation of month $t+1$ with its training-period climatological mean $\bar{P}_{m(t+1)}$:

$$SPI_k^{proxy}(t+1) = \Phi^{-1}[H_{m(t+1)}(\sum_{j=0}^{k-2} P(t-j) + \bar{P}_{m(t+1)})] \quad (3)$$

where $H_{m(t+1)}$ is the fitted distribution of the target calendar month. The learning models were trained to predict the residual $r(t+1) = SPI_k(t+1) - SPI_k^{proxy}(t+1)$, and their final forecast was the proxy plus the predicted residual. The input vector contained the proxy, SPI_k at lags 0–2, the known partial accumulation, the target-month climatology, precipitation at lags 0–1, T , SM , SLP and WS of month t , and the sine and cosine of the target month (15 inputs). The GRU received sequences of the previous 12 months of $[SPI_k, \text{proxy}, P, T, SM, SLP, WS, \sin, \cos]$.

2.4. Data Partitioning and Benchmarks

Forecasts were issued for target months from 2003 to 2023 and split chronologically by target month into training (to December 2016), validation (2017–2019) and test (2020–2023) periods. After removing the months needed for lags and for the GRU warm-up, the SPI-3 models used 163, 36 and 48 samples and the SPI-6 models 160, 36 and 48 samples, respectively. Two benchmarks were used: persistence ($SPI_k(t+1) = SPI_k(t)$) and the proxy of Eq. (3). All inputs were standardized with training-set statistics.

2.5. Learning Models

MLP. A feed-forward network with two hidden layers (16 and 8 ReLU neurons), trained by the Adam

optimizer with early stopping, was used as the standalone neural network (Rumelhart et al., 1986; Pedregosa et al., 2011).

XGBoost. Extreme gradient boosting builds an additive ensemble of regression trees f_k by minimizing a regularized objective (Chen and Guestrin, 2016):

$$L = \sum_i l(y_i, \hat{y}_i) + \sum_k [\gamma T_k + \frac{1}{2} \lambda \|w_k\|^2] \quad (4)$$

where l is the squared error, T_k is the number of leaves and w_k are the leaf weights of tree k . The standalone model used 300 trees, a maximum depth of 4 and a learning rate of 0.05.

GRU. The gated recurrent unit (Cho et al., 2014) is a simplified recurrent cell with update (z_t) and reset (r_t) gates:

$$z_t = \sigma(W_z x_t + U_z h_{t-1}), \quad r_t = \sigma(W_r x_t + U_r h_{t-1}) \quad (5)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tanh(W_h x_t + U_h (r_t \odot h_{t-1})) \quad (6)$$

The network had 24 GRU units, a dropout rate of 0.2 and a 12-unit dense layer. It was trained with Adam (learning rate 0.003, batch size 16) and early stopping on the validation loss (patience 40). The forecasts of five networks with different random seeds were averaged.

2.6. Metaheuristic Hybrid Models

PSO (Kennedy and Eberhart, 1995) moves a swarm of particles through the search space according to their own best position p_i and the global best position g :

$$v_i(t+1) = w v_i(t) + c_1 r_1 (p_i - x_i(t)) + c_2 r_2 (g - x_i(t)), \\ x_i(t+1) = x_i(t) + v_i(t+1) \quad (7)$$

GWO (Mirjalili et al., 2014) updates each wolf towards the three best solutions α , β and δ with coefficient vectors that shift the search from exploration to exploitation. **HHO** (Heidari et al., 2019) imitates the surprise pounce of Harris hawks. Depending on the escaping energy of the prey, $E = 2E_0(1 - t/T)$, the hawks switch between exploration and four besiege strategies, including Lévy-flight rapid dives.

In MLP-PSO the optimized hyperparameters were the sizes of the two hidden layers (4–64 and 2–32), the L2 penalty (10^{-5} – 10^{-1}) and the initial learning rate ($10^{-3.5}$ – $10^{-1.5}$). In XGB-GWO and XGB-HHO they were the number of trees (50–600), maximum depth (2–8), learning rate ($10^{-2.5}$ – $10^{-0.5}$), row and column subsampling (0.5–1), minimum child weight (1–10) and L2 regularization λ (10^{-2} – $10^{1.5}$). The fitness was the RMSE of the validation residuals of a model trained on the training set. Each optimizer used 15 agents and 30 iterations (MEALPY; Van Thieu and Mirjalili, 2023).

2.7. Evaluation

Continuous forecasts were evaluated with R^2 , RMSE, MAE, NSE (Nash and Sutcliffe, 1970) and Willmott's index (WI; Willmott, 1981), as well as Taylor diagrams (Taylor, 2001). Categorical skill was measured by the accuracy of the seven SPI classes, and for drought warnings ($SPI \leq -1$) by the probability of detection (POD) and the false alarm ratio (FAR; Wilks, 2011):

$$POD = H / (H + M), \quad FAR = F / (H + F) \quad (8)$$

where H , M and F are the numbers of hits, misses and false alarms.

3. Results

3.1. Drought Characteristics

The SPI series (Fig. 3) show alternating dry and wet phases. Moderate or worse drought ($SPI \leq -1$) occurred in 18%, 20% and 23% of the months for SPI-3, SPI-6 and SPI-12, respectively. The most notable dry spells were 2008, 2013–2014, 2017 and 2021–2023. The lowest SPI-6 (-3.04) was recorded in October 2017, and SPI-3 reached -2.43 in February 2023. The wet year 2018–2019 produced SPI-12 values above 4, which is outside the range of the training period and illustrates the sensitivity of the SPI to the calibration period. SPI-12 showed a weak but significant decreasing tendency (Mann–Kendall $p = 0.03$), whereas the trends of SPI-3 and SPI-6 were not significant. Most of the test period (2020–2023) was drier than normal: the mean test-period SPI-6 was -0.44, with 10 months below -1.

3.2. Hyperparameter Optimization

Table 2. Optimized hyperparameters and validation fitness (first → final iteration)

Index	Model	Selected hyperparameters	Fitness
SPI-3	MLP-PSO	layers = (52, 22), $\alpha = 2.60e-04$, lr = 0.0012	1.245 → 1.126
SPI-3	XGB-GWO	trees = 202, depth = 7, lr = 0.0217, $\lambda = 31.62$	1.300 → 1.289
SPI-3	XGB-HHO	trees = 292, depth = 2, lr = 0.0032, $\lambda = 0.01$	1.308 → 1.287
SPI-6	MLP-PSO	layers = (51, 24), $\alpha = 1.41e-02$, lr = 0.0118	1.210 → 1.141
SPI-6	XGB-GWO	trees = 560, depth = 2, lr = 0.0075, $\lambda = 1.42$	1.189 → 1.108
SPI-6	XGB-HHO	trees = 50, depth = 2, lr = 0.2061, $\lambda = 0.31$	1.234 → 1.064

The optimized configurations (Table 2) differed markedly between the two indices, which confirms that hyperparameters cannot simply be transferred between forecasting problems. For SPI-6 both XGBoost hybrids selected very shallow trees (depth 2), and for SPI-3 XGB-HHO selected shallow trees with a very small learning rate. Both choices correspond to strongly regularized, low-variance models. The convergence curves (Fig. 4) show that the optimizers improved the validation fitness by 1–14% within 30 iterations, with the largest gains for MLP-PSO (SPI-3) and XGB-HHO (SPI-6). HHO found the lowest validation fitness for SPI-6, and GWO and HHO performed equally for SPI-3.

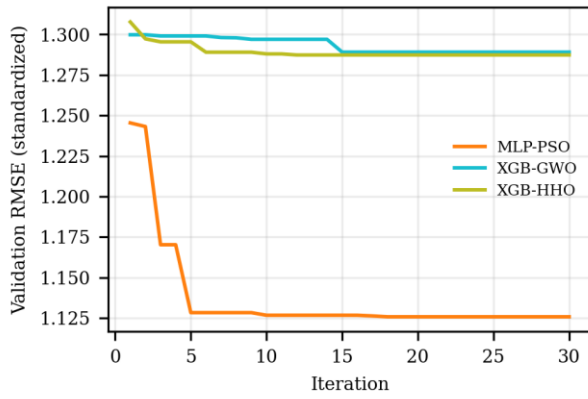


Figure 4a. Convergence of PSO, GWO and HHO for SPI-3 (validation RMSE of standardized residuals)

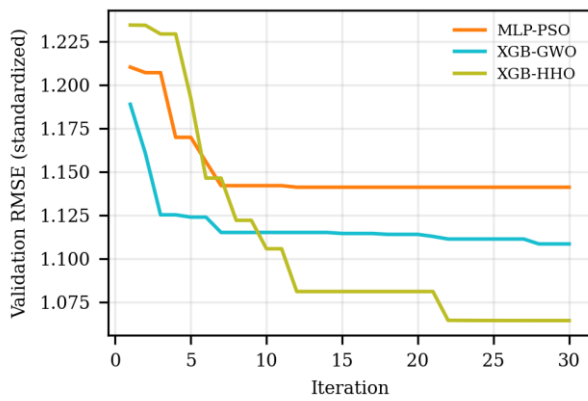


Figure 4b. Convergence of PSO, GWO and HHO for SPI-6

3.3. Forecast Accuracy

Table 3. Test-period (2020–2023) performance of one-month-ahead SPI forecasts

Model	R ² (SPI-3)	RMSE (SPI-3)	NSE (SPI-3)	R ² (SPI-6)	RMSE (SPI-6)	NSE (SPI-6)
Persistence	0.345	0.703	0.174	0.578	0.511	0.519
Proxy	0.685	0.458	0.650	0.787	0.360	0.761
MLP	0.475	0.684	0.219	0.331	0.842	-0.309
XGBoost	0.541	0.579	0.440	0.590	0.511	0.518
GRU	0.690	0.464	0.639	0.788	0.354	0.768
MLP-PSO	0.648	0.510	0.565	0.751	0.402	0.702
XGB-GWO	0.673	0.483	0.610	0.646	0.512	0.517
XGB-HHO	0.683	0.449	0.662	0.598	0.544	0.454

Training RMSE of the untuned XGBoost was 0.057 (SPI-3) and 0.030 (SPI-6), indicating strong overfitting.

Table 3 lists the test-period scores. Persistence was the weakest reference for both indices (SPI-3: RMSE = 0.703, NSE = 0.17; SPI-6: RMSE = 0.511, NSE = 0.52). The climatology-completed proxy, which needs no training, reduced the RMSE by 35% for SPI-3 and 29% for SPI-6 and reached NSE values of 0.65 and 0.76.

For SPI-3, XGB-HHO was the most accurate model (RMSE = 0.449, R² = 0.683, NSE = 0.662), followed by

the proxy (0.458), GRU (0.464), XGB-GWO (0.483) and MLP-PSO (0.510). For SPI-6, the GRU achieved the lowest error (RMSE = 0.354, NSE = 0.768), practically equal to the proxy (0.360), followed by MLP-PSO (0.402). The XGBoost-based models did not improve on the proxy for SPI-6. They had the lowest validation-period RMSE (0.49–0.51) but lost accuracy in the drier test period, which points to a shift in the residual distribution between periods.

The benefit of metaheuristic tuning was clearest when compared with the untuned learners. The standalone XGBoost almost reproduced the training data (training RMSE < 0.06) but generalized poorly, and the standalone MLP was the worst model for both indices (test RMSE 0.684 and 0.842). Tuning reduced the test RMSE of MLP by 25% (SPI-3) and 52% (SPI-6), and that of XGBoost by 22% for SPI-3 with HHO.

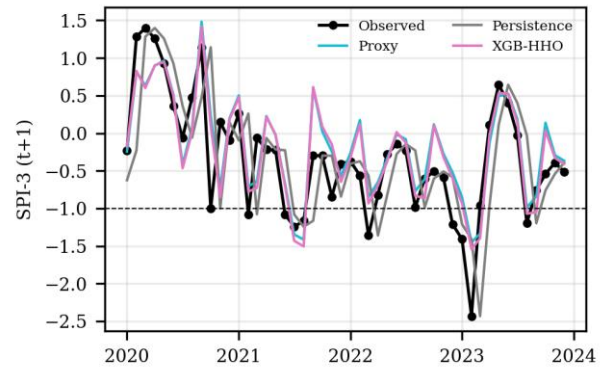


Figure 5a. Observed and forecast SPI-3 in the test period

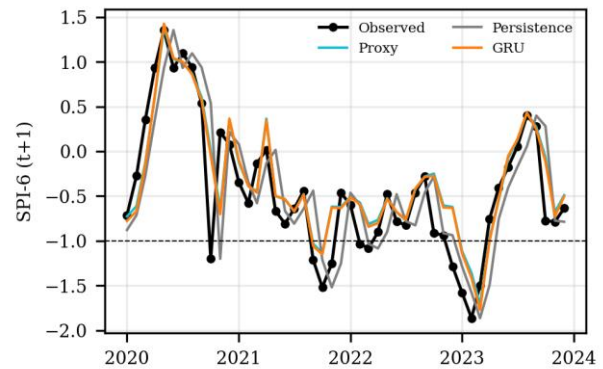


Figure 5b. Observed and forecast SPI-6 in the test period

The test-period time series (Fig. 5) show that the proxy and the best learners followed the observed evolution of the 2021–2023 drought closely, while persistence systematically lagged by one month. Errors were largest at sharp transitions caused by an unexpectedly wet or dry month, such as the onset of drought in late 2020 and the partial recovery in spring 2023. These events cannot be anticipated from antecedent information alone. The scatter plots and Taylor diagrams (Figs. 6 and 7) confirm this picture. The GRU, the proxy and MLP-PSO had correlations of 0.81–0.89 with normalized standard deviations of 0.84–1.00,

while the tuned XGBoost models reached 0.77–0.83. The untuned MLP had the lowest correlation (0.69 and 0.58).

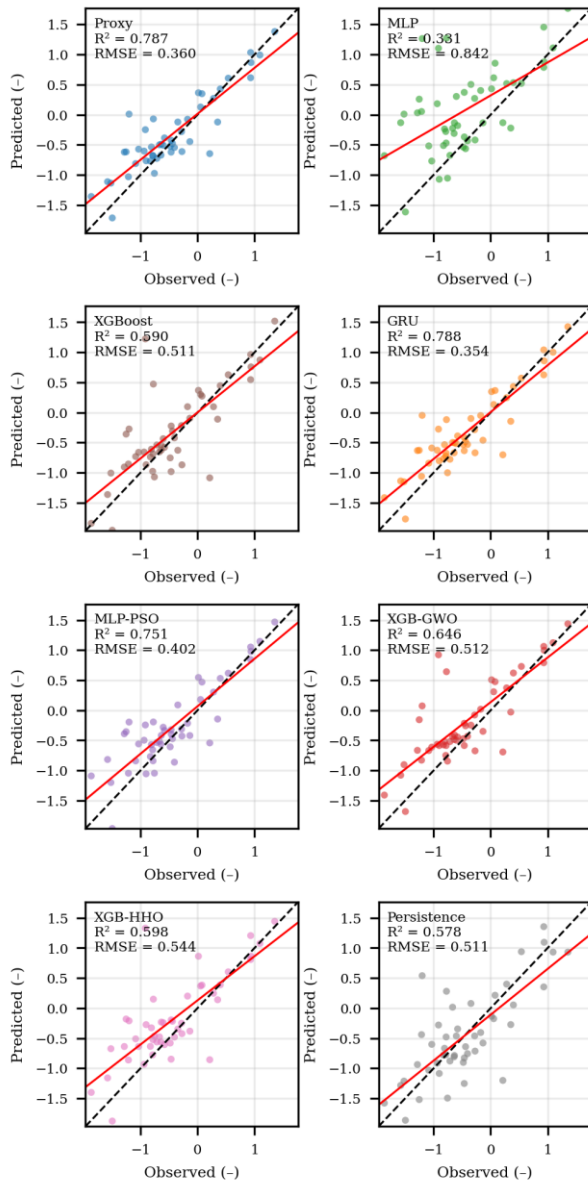


Figure 6. Observed versus forecast SPI-6 in the test period

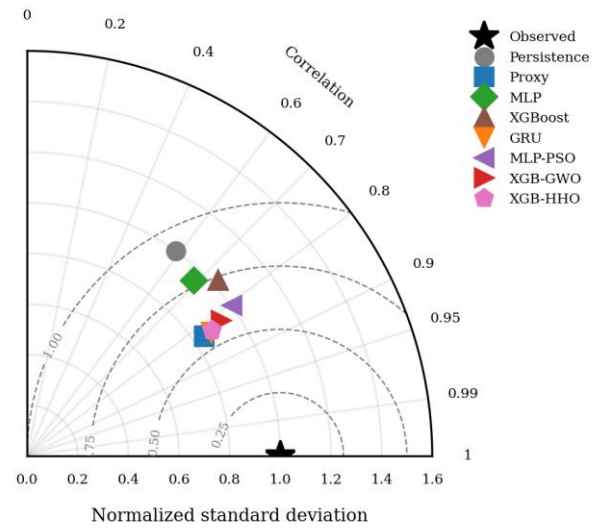


Figure 7a. Taylor diagram of SPI-3 forecasts (test period)

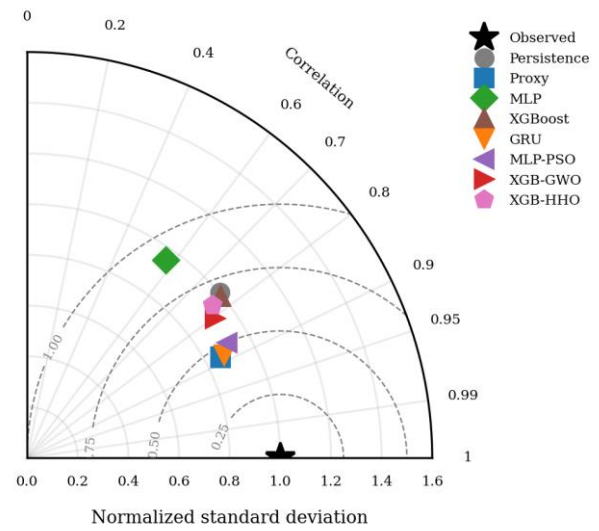


Figure 7b. Taylor diagram of SPI-6 forecasts (test period)

3.4. Drought-Class Skill

Table 4. Categorical skill in the test period (accuracy of SPI classes and drought warnings for $SPI \leq -1$)

Model	Acc. % (SPI-3)	POD (SPI-3)	FAR (SPI-3)	Acc. % (SPI-6)	POD (SPI-6)	FAR (SPI-6)
Persistence	69	0.44	0.56	67	0.60	0.40
Proxy	77	0.33	0.25	77	0.50	0.00
MLP	73	0.33	0.25	67	0.10	0.67
XGBoost	73	0.56	0.44	71	0.40	0.33
GRU	75	0.33	0.25	77	0.50	0.17
MLP-PSO	81	0.67	0.33	75	0.40	0.33
XGB-GWO	71	0.33	0.25	77	0.30	0.00
XGB-HHO	75	0.44	0.33	75	0.30	0.00

The test period contained 9 (SPI-3) and 10 (SPI-6) drought months.

With only 9–10 drought months in the test period, the categorical scores (Table 4) must be interpreted with

care. MLP-PSO achieved the highest class accuracy (81%) and POD (0.67) for SPI-3. For SPI-6, the proxy, GRU and XGB-GWO reached the highest accuracy (77%), and the proxy combined the highest POD (0.50) with no false alarms. Persistence detected drought onset one month late and had one of the highest false alarm ratios, mainly at drought termination.

3.5. Predictor Importance

For SPI-3 the permutation importance of the XGBoost residual model (Fig. 8) was dominated by P(t) and SPI3(t-2), followed by SM(t). These variables describe the part of the window that leaves the accumulation and the recent wetness state. For SPI-6 the most important inputs were the SPI6-proxy, SPI6(t), P(t) and SPI6(t-1). Temperature, pressure and wind speed contributed little, which is consistent with the weak monthly predictability of precipitation in the region from local atmospheric variables.

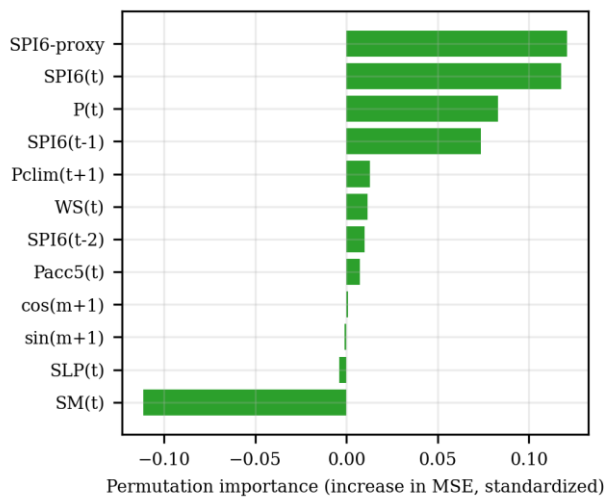


Figure 8. Permutation importance of the inputs of the XGBoost residual model for SPI-6

4. Discussion

The most important finding of this study is the strength of the climatology-completed proxy. Most of the predictability of SPI one month ahead comes from the overlap of the accumulation windows. This is especially true for longer windows, since five of the six months of SPI-6(t+1) are already known. Many published SPI forecasting studies report very high NSE values for SPI-6 and SPI-12 (e.g. Çıtakoğlu and Coşkun, 2022; Elbeltagi et al., 2023), but without a proxy benchmark it is difficult to know how much of this skill is contributed by the learner. In our case, the best learners improved on the proxy by only 2% (SPI-3) and 1.6% (SPI-6) in RMSE. This is consistent with the limited predictability of monthly precipitation in semi-arid regions (Hao et al., 2018).

Second, the value of metaheuristic optimization was mainly a regularization effect. Untuned XGBoost and MLP overfitted the 160 training samples, and the optimizers moved them towards simpler configurations:

shallow trees, slow learning, stronger penalties and larger but better-regularized networks. The HHO, GWO and PSO searches were computationally cheap (10–40 s) and therefore worthwhile. However, the model with the lowest validation fitness was not always the best in the test period (e.g. XGB-HHO for SPI-6), so validation-based selection over small samples remains uncertain. Using the validation period both for early stopping and for hyperparameter search also makes the validation scores optimistic, and only the test scores should be used for comparison.

Third, fitting the gamma distribution on the training period only is necessary for an honest evaluation, but it has a visible side effect. The very wet year 2018–2019 produced SPI values above 4, which would not occur if the whole record were used. Longer station records (≥ 30 years), multi-station analysis and large-scale climate predictors such as ENSO and NAO indices, as well as precipitation forecasts from seasonal prediction systems, are likely to add more skill than further tuning of the learners. Probabilistic forecasts would also be more useful for drought early warning than deterministic ones.

5. Conclusions

One-month-ahead forecasts of SPI-3 and SPI-6 were developed for the Lake Urmia basin using MLP, XGBoost, GRU and three metaheuristic hybrids trained as residual corrections of a climatology-completed SPI proxy. The main conclusions are: (1) the proxy, which requires no training, reduced the RMSE of persistence by 35–29% and should be used as a standard benchmark; (2) XGB-HHO (SPI-3, RMSE = 0.449) and GRU (SPI-6, RMSE = 0.354) were the most accurate models, but their gain over the proxy was small; (3) PSO, GWO and HHO effectively controlled the overfitting of MLP and XGBoost and reduced their test errors by up to 52%; and (4) the basin experienced persistent drought during 2020–2023, which the best models tracked with a class accuracy of 75–81%. Physics-informed inputs and honest calibration are therefore more important than the choice of learner for short-lead SPI forecasting.

6. Data Availability

Monthly IRIMO climate data and GLDAS soil moisture for the Lake Urmia basin were obtained from github.com/habibisara/Urmia-Lake-Climate-Study (Habibi and Tasouji Hassanpour, 2025). The code used for the analysis is available from the authors on request.

7. References

- AghaKouchak, A., Norouzi, H., Madani, K., Mirchi, A., Azarderakhsh, M., Nazemi, A., Nasrollahi, N., Farahmand, A., Mehran, A., & Hasanzadeh, E. (2015). Aral Sea syndrome desiccates Lake Urmia: Call for action. *Journal of Great Lakes Research*, 41(1), 307–311.
- Belayneh, A., Adamowski, J., Khalil, B., & Ozga-Zielinski, B. (2014). Long-term SPI drought forecasting in the

- Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. *Journal of Hydrology*, 508, 418–429.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794).
- Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of EMNLP 2014* (pp. 1724–1734).
- Çitakoğlu, H., & Coşkun, Ö. (2022). Comparison of hybrid machine learning methods for the prediction of short-term meteorological droughts of Sakarya Meteorological Station in Turkey. *Environmental Science and Pollution Research*, 29. <https://doi.org/10.1007/s11356-022-21083-3>
- Elbeltagi, A., et al. (2023). Prediction of meteorological drought and standardized precipitation index based on the random forest (RF), random tree (RT), and Gaussian process regression (GPR) models. *Environmental Science and Pollution Research*, 30.
- Habibi, S., & Tasouji Hassanpour, S. (2025). An explainable machine learning framework for forecasting lake water equivalent using satellite data: A 20-year analysis of the Urmia Lake Basin. *Water*, 17(10), 1431.
- Hao, Z., Singh, V. P., & Xia, Y. (2018). Seasonal drought prediction: Advances, challenges, and future prospects. *Reviews of Geophysics*, 56(1), 108–141.
- Heidari, A. A., Mirjalili, S., Faris, H., Aljarah, I., Mafarja, M., & Chen, H. (2019). Harris hawks optimization: Algorithm and applications. *Future Generation Computer Systems*, 97, 849–872.
- Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95 – International Conference on Neural Networks* (Vol. 4, pp. 1942–1948). IEEE.
- McKee, T. B., Doesken, N. J., & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology* (pp. 179–184). American Meteorological Society, Anaheim, CA.
- Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. *Advances in Engineering Software*, 69, 46–61.
- Mishra, A. K., & Desai, V. R. (2006). Drought forecasting using feed-forward recursive neural network. *Ecological Modelling*, 198(1–2), 127–138.
- Mishra, A. K., & Singh, V. P. (2010). A review of drought concepts. *Journal of Hydrology*, 391(1–2), 202–216.
- Moghaddasi, M., et al. (2025). Reinforcing long lead time drought forecasting with a novel hybrid deep learning model: A case study in Iran. *Applied Water Science*.
- Mohamadi, S., et al. (2020). Zoning map for drought prediction using integrated machine learning models with a nomadic people optimization algorithm. *Natural Hazards*, 104(1). <https://doi.org/10.1007/s11069-020-04180-9>
- Nabipour, N., Dehghani, M., Mosavi, A., & Shamshirband, S. (2020). Short-term hydrological drought forecasting based on different nature-inspired optimization algorithms hybridized with artificial neural networks. *IEEE Access*, 8.
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3), 282–290.
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Piri, J., Abdolahipour, M., et al. (2023). Advanced machine learning model for prediction of drought indices using hybrid SVR-RSM. *Water Resources Management*, 37(2). <https://doi.org/10.1007/s11269-022-03395-8>
- Raziei, T., Saghafian, B., Paulo, A. A., Pereira, L. S., & Bordini, I. (2009). Spatial patterns and temporal variability of drought in western Iran. *Water Resources Management*, 23, 439–455.
- Rodell, M., Houser, P. R., Jambor, U., et al. (2004). The Global Land Data Assimilation System. *Bulletin of the American Meteorological Society*, 85(3), 381–394.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533–536.
- Stagge, J. H., Tallaksen, L. M., Gudmundsson, L., Van Loon, A. F., & Stahl, K. (2015). Candidate distributions for climatological drought indices (SPI and SPEI). *International Journal of Climatology*, 35(13), 4027–4040.
- Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, 106(D7), 7183–7192.
- Van Thieu, N., & Mirjalili, S. (2023). MEALPY: An open-source library for latest meta-heuristic algorithms in Python. *Journal of Systems Architecture*, 139, 102871.
- Wilks, D. S. (2011). *Statistical Methods in the Atmospheric Sciences* (3rd ed.). Academic Press, Oxford.
- Willmott, C. J. (1981). On the validation of models. *Physical Geography*, 2(2), 184–194.
- World Meteorological Organization (WMO). (2012). *Standardized Precipitation Index User Guide* (WMO-No. 1090). Geneva.