

Forecasting Terrestrial Water Storage Anomalies in the Lake Urmia Basin Using Machine Learning, LSTM and Metaheuristic-Optimized Support Vector Regression

Hatice Citakoglu¹

¹ ERCIYES UNIVERSITY Department of Civil Engineering (✉ hcitakoglu@erciyes.edu.tr)

Received: 15 September 2025

Accepted: 17 October 2025

Available online: 19 December 2025

Abstract: Lake Urmia in north-western Iran has experienced one of the most severe hydrological declines of any large saline lake in the world, and the terrestrial water storage (TWS) of its basin is a key indicator for water-resources planning. This study evaluates classical machine learning (multiple linear regression, support vector regression and random forest), a deep learning model (long short-term memory, LSTM) and two hybrid models in which the hyperparameters of support vector regression were tuned by the Grey Wolf Optimizer (SVR-GWO) and the Whale Optimization Algorithm (SVR-WOA) for one-month-ahead estimation of GRACE-derived liquid water equivalent (LWE) anomalies. Monthly basin-average LWE (GRACE, 2003–2017) together with precipitation, air temperature, sea-level pressure and wind speed from the Iran Meteorological Organization and GLDAS soil moisture were used. The LWE series showed a highly significant decreasing trend (Mann–Kendall $Z = -11.57$, Sen's slope = -2.02 cm/year). All models were trained on 2004–2012, tuned on 2013–2014 and tested on 2015–2017 and were compared against a persistence benchmark. In the test period the LSTM gave the lowest error (RMSE = 3.16 cm, NSE = 0.76), followed closely by MLR (RMSE = 3.19 cm), SVR-WOA (3.30 cm) and SVR-GWO (3.31 cm), while the persistence benchmark reached 4.17 cm. Metaheuristic tuning reduced the test RMSE of SVR by about 2%, and all learning models improved on persistence by 17–24%. Soil moisture and air temperature were the most influential predictors. The small differences among the best models show that, for short and strongly seasonal GRACE records, simple well-regularized models remain competitive with deep learning.

Keywords: Lake Urmia, GRACE, Terrestrial water storage, LSTM, Grey Wolf Optimizer, Whale Optimization Algorithm

1. Introduction

Lake Urmia, located in the north-west of Iran, was once the largest lake in the Middle East and one of the largest hypersaline lakes on Earth. Since the late 1990s its water level has fallen by several metres and its surface area has shrunk dramatically, a process that has been attributed to a combination of prolonged droughts, dam construction, expansion of irrigated agriculture and over-abstraction of groundwater (AghaKouchak et al., 2015; Tourian et al., 2015). The desiccation has created salt-dust hazards and threatens agriculture, public health and the ecology of the whole basin. Reliable information on how much water is stored in the basin, and how this storage is likely to evolve in the coming months, is therefore essential for operational water allocation and for evaluating the national Lake Urmia restoration programme.

Ground-based monitoring alone cannot describe the total amount of water stored in soils, aquifers, reservoirs and the lake itself. The Gravity Recovery and Climate Experiment (GRACE) mission and its successor GRACE Follow-On have provided monthly estimates of terrestrial water storage anomalies (TWSA) since 2002 (Tapley et al., 2004; Landarer et al., 2020). GRACE data have been used extensively in the Middle East to reveal large-scale groundwater depletion (Voss et al., 2013; Joodaki et al., 2014), and several studies have analysed the Lake Urmia basin in particular (Khorrami et al., 2023; Sabzehee et al., 2023). However, GRACE products are released with a latency of several months,

the record contains gaps, and the coarse resolution complicates their direct use in local decision making. Data-driven models that relate TWSA to readily available hydro-climatic variables are therefore attractive both for gap filling and for short-term forecasting (Sun et al., 2020).

Machine learning (ML) has become a standard tool for this purpose. Soltani et al. (2022, 2024) modelled TWSA of the Lake Urmia basin with the group method of data handling, adaptive neuro-fuzzy inference systems, extreme learning machines and artificial neural networks using ERA5 and GLDAS predictors. Radman et al. (2022) compared feed-forward and recurrent deep learning architectures to predict the surface area of Lake Urmia and found long short-term memory (LSTM) networks to be the most accurate. Foroumandi et al. (2023) downscaled GRACE TWSA over Iran with convolutional LSTM networks, and Habibi and Tasouji Hassanpour (2025) proposed an explainable random forest framework for forecasting lake water equivalent in the Urmia basin. These studies consistently show that non-linear learners can reproduce storage dynamics, but they also reveal that model performance depends strongly on the choice of hyperparameters.

Metaheuristic optimization algorithms such as the Grey Wolf Optimizer (GWO; Mirjalili et al., 2014) and the Whale Optimization Algorithm (WOA; Mirjalili and Lewis, 2016) provide a derivative-free way to tune these hyperparameters. Hybrid models built in this way have been reported to outperform their standalone

counterparts in groundwater-level prediction (Seifi et al., 2020; Kayhomayoon et al., 2022; Mozaffari et al., 2022). In many published comparisons, however, the hybrid model is compared only with an untuned version of the same learner and not with simple benchmarks, and the hyperparameters are sometimes selected using the test data. According to the no-free-lunch theorem (Wolpert and Macready, 1997), no single optimizer or learner can be expected to be superior for every problem, so the benefit of hybridization should be demonstrated with an independent test period and a naive reference forecast.

The objectives of this study were therefore (i) to characterize the long-term behaviour of GRACE-derived water storage in the Lake Urmia basin, (ii) to develop one-month-ahead estimation models of LWE anomalies using multiple linear regression (MLR), support vector regression (SVR), random forest (RF), LSTM and two hybrid models (SVR-GWO and SVR-WOA), and (iii) to assess these models against a persistence benchmark in a strictly chronological train–validation–test design, identifying the dominant hydro-climatic drivers of storage variability.

2. Materials and Methods

2.1. Study Area

The Lake Urmia basin (Fig. 1) lies between 44°13' and 47°53' E and 35°40' and 38°30' N and covers about 51,760 km² in the provinces of West Azerbaijan, East Azerbaijan and Kurdistan (Habibi and Tasouji Hassanpour, 2025). The climate is semi-arid and continental, with cold winters, dry summers and a mean annual basin precipitation of roughly 300–350 mm, most of which falls between October and May. More than a dozen permanent rivers drain towards the terminal lake, and a large part of their flow is diverted for irrigation. The basin therefore combines strong natural seasonality with a pronounced anthropogenic signal.

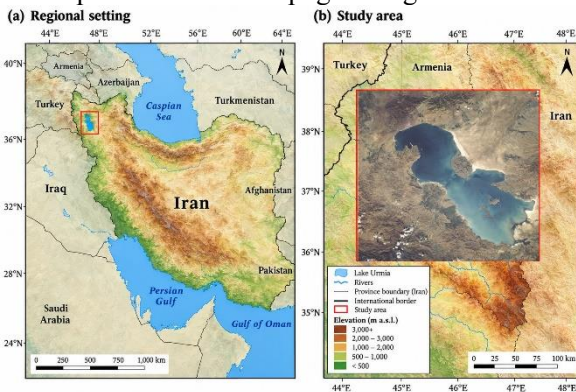


Figure 1. Location of the Lake Urmia basin in north-western Iran (the rectangle shows the approximate bounding box of the basin)

2.2. Data

The monthly dataset used in this study was obtained from the public repository of Habibi and Tasouji Hassanpour (2025) ([github.com/habibisara/Urmia-](https://github.com/habibisara/Urmia-Lake-Climate-Study)

Lake-Climate-Study), which compiles basin-averaged variables for January 2003 to December 2023. Liquid water equivalent (LWE, cm) anomalies were derived from GRACE (2003–2017) and GRACE Follow-On (2018–2023) mascon products in Google Earth Engine. Air temperature (T, °C), precipitation (P, mm), sea-level pressure (SLP, reported in Pa and converted to hPa) and wind speed (WS, m/s) were taken from the Iran Meteorological Organization (IRIMO), and volumetric soil moisture (SM, %) was taken from the Global Land Data Assimilation System (GLDAS; Rodell et al., 2004). Months missing from the GRACE record had been linearly interpolated by the data providers.

A preliminary inspection showed that from mid-2018 to early 2023 the shared LWE series has a seasonal amplitude of only 1–2 cm, compared with 10–20 cm during the GRACE period, followed by an abrupt drop to about –20 cm in 2023. Because this change coincides with the transition between the two satellite missions and is not reflected in the meteorological drivers, it most likely reflects processing differences rather than hydrology. To avoid training and evaluating models on an inhomogeneous target, the analysis was restricted to the GRACE mission period (January 2003 to June 2017, 174 months). The descriptive statistics of the variables are given in Table 1 and their time series are shown in Fig. 2.

Table 1. Descriptive statistics of the monthly variables (January 2003 – June 2017)

Variable	Min	Max	Mean	SD	Skew	Kurt
LWE (cm)	-28.26	14.30	-6.80	10.50	0.05	-0.79
P (mm)	0.37	124.51	28.01	21.95	1.00	1.34
T (°C)	-9.59	23.81	8.50	9.68	-0.02	-1.36
SM (%)	12.52	42.15	24.44	7.23	0.10	-0.81
SLP (hPa)	829.79	843.07	836.10	2.60	0.25	-0.34
WS (m/s)	1.57	4.14	2.60	0.45	0.16	0.32

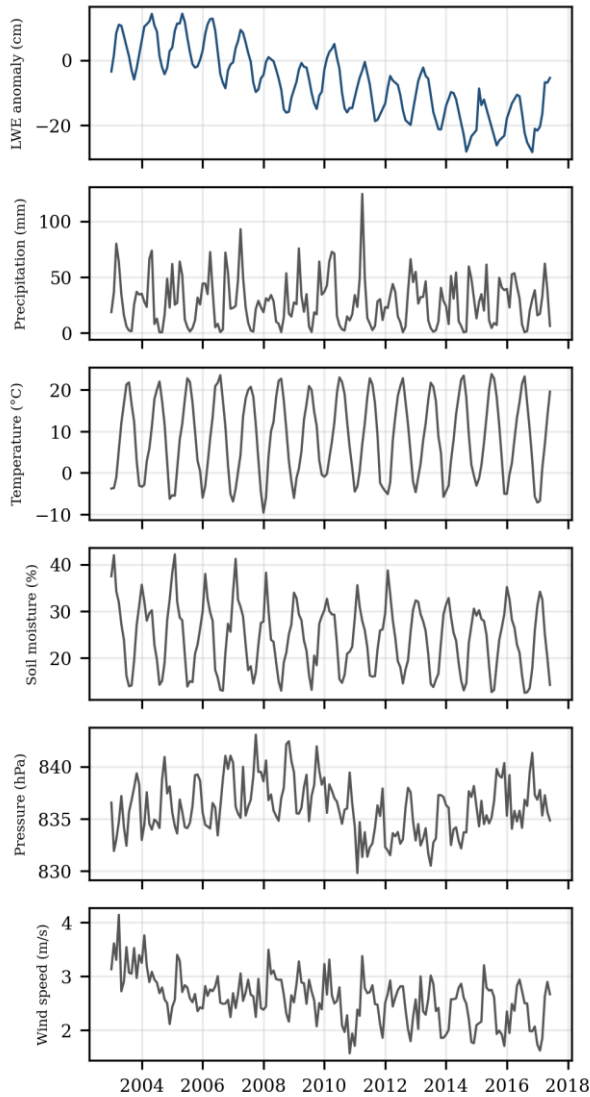


Figure 2. Monthly time series of LWE anomaly and the hydro-climatic predictors in the Lake Urmia basin

The LWE anomaly declined from about +5 cm in 2004–2005 to about –18 cm in 2014–2016, reaching its minimum of –28.3 cm in November 2016. The non-parametric Mann–Kendall test (Mann, 1945; Kendall, 1975) with Sen's (1968) slope estimator confirmed a significant decreasing trend ($Z = -11.57$, $p < 0.001$, slope = -0.168 cm/month, i.e. -2.02 cm/year). The Pearson correlation matrix (Fig. 3) shows that, at the same month, LWE is only weakly related to precipitation ($r = 0.24$) and soil moisture ($r = 0.31$). The strongest same-month correlation was with wind speed ($r = 0.60$), which mainly reflects the shared seasonal cycle, while soil moisture and temperature were strongly anti-correlated ($r = -0.86$).

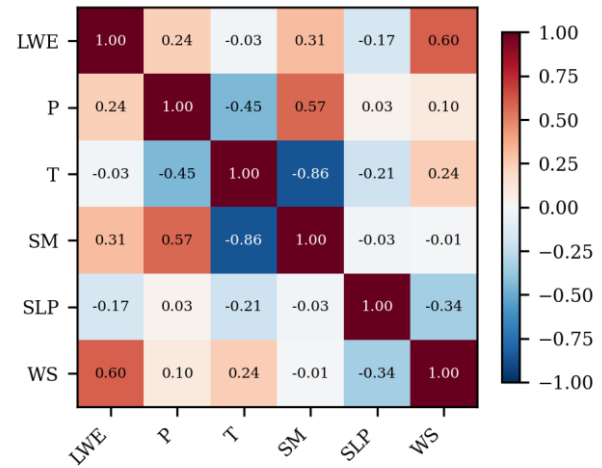


Figure 3. Pearson correlation matrix of the monthly variables

2.3. Model Inputs and Data Partitioning

The target was the LWE anomaly of month t . The candidate predictors were the LWE anomalies of the three preceding months, precipitation at lags 0–2, temperature and soil moisture at lags 0–1, SLP and WS of month t , and a sine–cosine encoding of the calendar month. Because the lagged storage dominates the variance of a strongly autocorrelated and trending series, all models were trained to predict the monthly increment $\Delta LWE(t) = LWE(t) - LWE(t-1)$, and the forecast of the level was obtained by adding the increment to the last observed value. This formulation avoids extrapolation problems of tree-based learners when the test period lies outside the range of the training data. For the LSTM, sequences of the previous 12 months of [LWE($t-1$), P, T, SM, SLP, WS, sin, cos] were used.

The data were split chronologically into training (April 2004 – December 2012, $n = 117$), validation (January 2013 – December 2014, $n = 24$) and test (January 2015 – June 2017, $n = 30$) sets. The first twelve months were used as a warm-up for the LSTM sequences, so that all models were evaluated on exactly the same months. All inputs were standardized with the mean and standard deviation of the training set only. The validation set was used for early stopping of the LSTM and as the objective of the metaheuristic search; the test set was used only once, for the final evaluation. A persistence forecast ($LWE(t) = LWE(t-1)$) was used as the naive benchmark.

2.4. Machine Learning Models

Multiple linear regression (MLR) was used as a transparent baseline, fitted by ordinary least squares on the standardized predictors.

Support vector regression (SVR) maps the inputs into a high-dimensional feature space with a kernel function and fits a function that deviates from the targets by at most ϵ while remaining as flat as possible (Drucker et al., 1997). The prediction is

$$f(x) = \sum (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (1)$$

where α_i and α_i^* are Lagrange multipliers and K is the radial basis function kernel $K(x_i, x) = \exp(-\gamma\|x_i - x\|^2)$. The penalty parameter C , the kernel width γ and the tube width ε control the bias–variance trade-off. The standalone SVR used the default values of the scikit-learn library ($C = 1$, $\varepsilon = 0.1$, $\gamma = \text{'scale'}$; Pedregosa et al., 2011).

Random forest (RF) is an ensemble of decision trees grown on bootstrap samples with random feature subsets, whose predictions are averaged (Breiman, 2001). Five hundred trees with a minimum of two samples per leaf were used. Permutation importance on the test set was used to rank the predictors.

2.5. Long Short-Term Memory Network

LSTM networks are recurrent neural networks whose memory cell is controlled by input, forget and output gates, which allows them to learn long-range temporal dependencies (Hochreiter and Schmidhuber, 1997). They have become a standard deep learning tool in hydrology (Kratzert et al., 2018). The cell state c_t and hidden state h_t are updated as

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f), \quad i_t = \sigma(W_i[h_{t-1}, x_t] + b_i), \quad o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (2)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c[h_{t-1}, x_t] + b_c), \quad h_t = o_t \odot \tanh(c_t) \quad (3)$$

The network used here had one LSTM layer with 32 units, a dropout rate of 0.2, a dense layer with 16 ReLU units and a linear output. It was trained with the Adam optimizer (learning rate 0.003, batch size 16, mean squared error loss) for up to 400 epochs with early stopping (patience 40) on the validation loss. To reduce the effect of random initialization, five networks with different seeds were trained and their predictions averaged. The models were implemented in Keras 3 with the JAX backend.

2.6. Hybrid SVR Models with GWO and WOA

GWO mimics the leadership hierarchy and hunting behaviour of grey wolves. The three best solutions (α , β and δ) guide the position update of the other wolves (Mirjalili et al., 2014):

$$D = |C \cdot X_p(t) - X(t)|, \quad X(t+1) = X_p(t) - A \cdot D \quad (4)$$

where X_p is the position of a leader, $A = 2a \cdot r_1 - a$ and $C = 2r_2$, with a decreasing linearly from 2 to 0 and r_1, r_2 random vectors in $[0, 1]$. The new position is the average of the moves towards α , β and δ . WOA simulates the bubble-net feeding of humpback whales and alternates between encircling the best solution and a logarithmic spiral movement (Mirjalili and Lewis, 2016):

$$X(t+1) = D' \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t) \quad (5)$$

where $D' = |X^*(t) - X(t)|$ is the distance to the best solution, b defines the spiral shape and l is a random number in $[-1, 1]$; the encircling and spiral moves are chosen with equal probability. In both hybrids each search agent encoded $\log_{10}C \in [-1, 3]$, $\log_{10}\gamma \in [-4, 0]$ and $\log_{10}\varepsilon \in [-3, -0.5]$. The fitness was the RMSE of the standardized validation-set predictions of an SVR trained on the training set. Both algorithms were run

with 20 agents for 40 iterations using the MEALPY library (Van Thieu and Mirjalili, 2023), and the final SVR was refitted on the training set with the best hyperparameters.

2.7. Evaluation Criteria

Model performance was evaluated with the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), Nash–Sutcliffe efficiency (NSE; Nash and Sutcliffe, 1970) and Willmott's index of agreement (WI; Willmott, 1981):

$$RMSE = \sqrt{[(1/n) \sum (O_i - P_i)^2]} \quad (6)$$

$$MAE = (1/n) \sum |O_i - P_i| \quad (7)$$

$$NSE = 1 - \sum (O_i - P_i)^2 / \sum (O_i - \bar{O})^2 \quad (8)$$

$$WI = 1 - \sum (O_i - P_i)^2 / \sum (|P_i - \bar{O}| + |O_i - \bar{O}|)^2 \quad (9)$$

where O_i and P_i are observed and predicted values and $\bar{O}$ is the observed mean. Taylor diagrams (Taylor, 2001) were used to summarize correlation, standard deviation and centred RMS difference in a single plot.

3. Results

3.1. Hyperparameter Optimization

The convergence curves of the two metaheuristics are shown in Fig. 4. Both algorithms reached their final fitness within about 20 iterations and then stabilized. GWO reached a slightly lower final fitness (0.3407) than WOA (0.3432). Both optimizers selected a large penalty parameter together with a very small kernel width (Table 2), i.e. an almost linear and weakly regularized SVR. This is consistent with the good performance of MLR reported below. Each optimization took less than 2 s on a standard CPU.

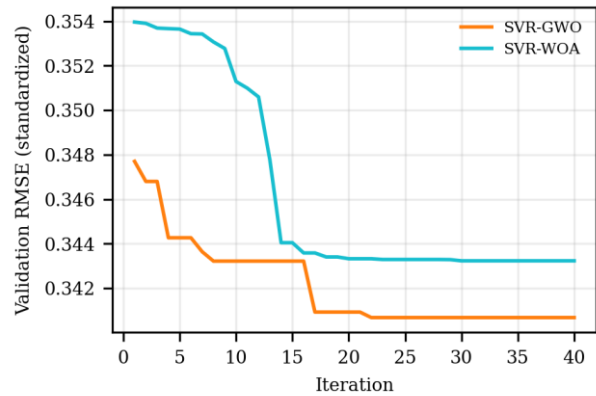


Figure 4. Convergence of the GWO and WOA searches (validation RMSE of standardized ΔLWE)

Table 2. SVR hyperparameters and validation fitness of the hybrid models

Model	C	γ	ε	Fitness
SVR (default)	1.00	scale	0.1000	–
SVR-GWO	554.10	0.00010	0.1525	0.3407
SVR-WOA	64.25	0.00039	0.1167	0.3432

3.2. Model Performance

Table 3. Performance of the models in the training and test periods (RMSE in cm)

Model	R ² (train)	RMSE (train)	NSE (train)	R ² (test)	RMSE (test)	NSE (test)
Persistence	0.865	3.336	0.860	0.610	4.166	0.576
MLR	0.983	1.176	0.983	0.762	3.188	0.751
SVR	0.987	1.002	0.987	0.732	3.373	0.722
RF	0.995	0.646	0.995	0.726	3.464	0.707
LSTM	0.983	1.175	0.983	0.761	3.157	0.756
SVR-GWO	0.980	1.275	0.980	0.748	3.307	0.733
SVR-WOA	0.979	1.298	0.979	0.746	3.305	0.733

Validation-period RMSE: MLR 1.25, SVR 1.60, RF 1.37, LSTM 1.24, SVR-GWO 1.13, SVR-WOA 1.14 and persistence 3.16 cm.

Table 3 summarizes the accuracy of all models. During training all learners fitted the data very closely ($R^2 \geq 0.98$), with RF showing the smallest training error (0.65 cm). This is a typical sign of the high flexibility of tree ensembles. In the independent test period (2015 – mid-2017), the ranking changed. The LSTM gave the lowest RMSE (3.16 cm) and MAE (2.12 cm) and the highest NSE (0.756) and WI (0.931). It was followed by MLR (RMSE = 3.19 cm), SVR-WOA (3.30 cm), SVR-GWO (3.31 cm), the default SVR (3.37 cm) and RF (3.46 cm). The persistence benchmark reached an RMSE of 4.17 cm and an NSE of 0.576. The learning models therefore reduced the forecast error of persistence by 17% (RF) to 24% (LSTM).

Hyperparameter tuning improved SVR in both validation and test periods. The test RMSE fell from 3.37 cm to 3.31 cm with GWO and to 3.30 cm with WOA, and R^2 increased from 0.732 to about 0.748. The two optimizers gave practically identical test scores, although they converged to different values of C and γ . This indicates a flat region of the fitness landscape rather than a single sharp optimum. The difference between the best (LSTM) and the fourth-ranked model (SVR-GWO) was only 0.15 cm, which is small compared with the month-to-month variability of LWE and with the uncertainty of GRACE solutions.

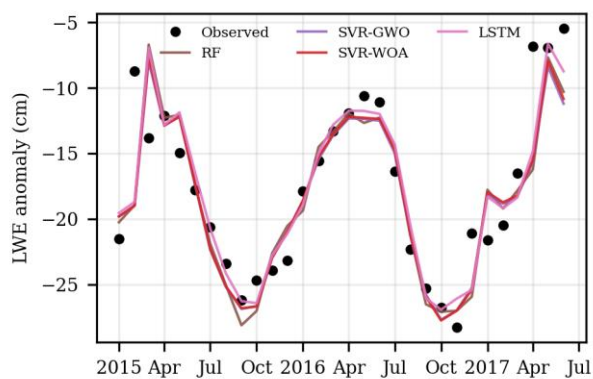


Figure 5. Observed and predicted LWE anomalies in the test period (2015 – June 2017)

The test-period hydrographs (Fig. 5) show that all models reproduced the seasonal cycle and the persistently low storage of 2015–2017. The largest errors occurred at sudden changes, such as the sharp rise in February–March 2015 and the recovery in early 2017, which the models captured only one month late. Fig. 6 shows the LSTM over the full record. The fit is tight in the training and validation periods, while in the test period the timing of rapid rises is sometimes missed by one month.

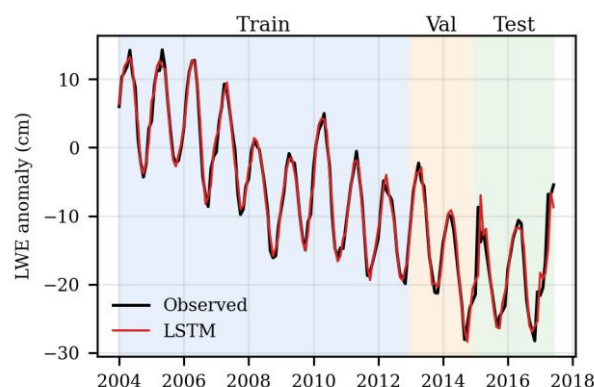


Figure 6. Observed and LSTM-predicted LWE anomalies for the training, validation and test periods

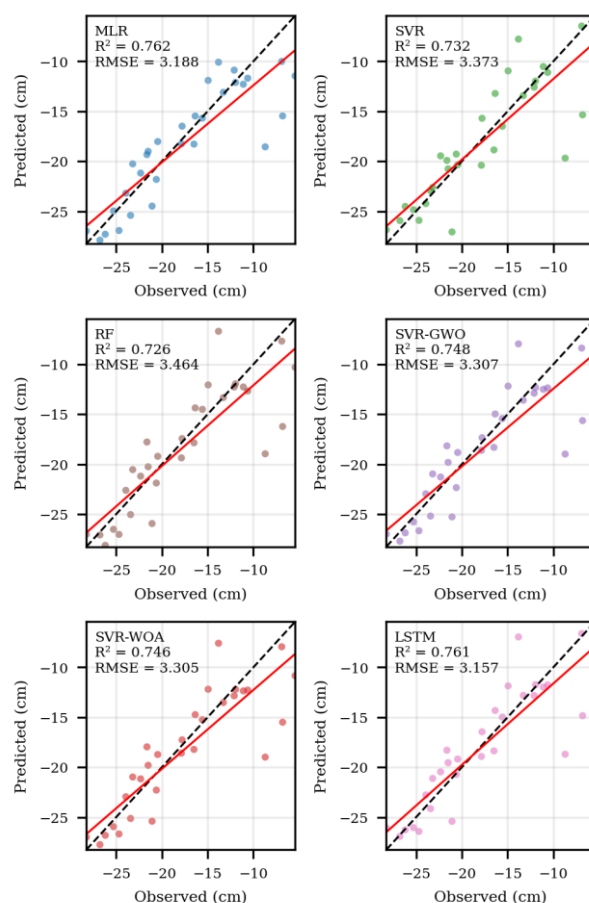


Figure 7. Scatter plots of observed versus predicted LWE in the test period (dashed: 1:1 line; red: regression line)

The scatter plots (Fig. 7) and the Taylor diagram (Fig. 8) show the same picture. All learning models lie close to each other, with correlation coefficients of 0.85–0.87 and normalized standard deviations of 0.88–0.95. Persistence has a similar spread but a lower correlation (0.78). The models therefore slightly underestimate the variance of the observations, a common effect of minimizing squared errors. The mean errors of all models were small and negative (−0.2 to −0.8 cm), i.e. storage was slightly underestimated. The interquartile range of the errors (Fig. 9) was 2.2–3.0 cm for the learning models and 4.5 cm for persistence.

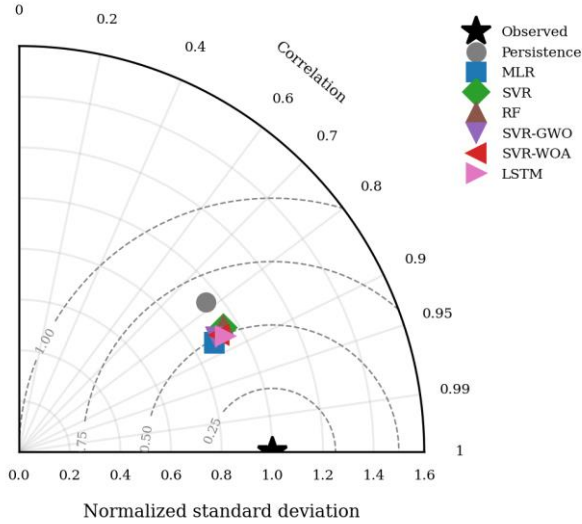


Figure 8. Taylor diagram of the test-period predictions

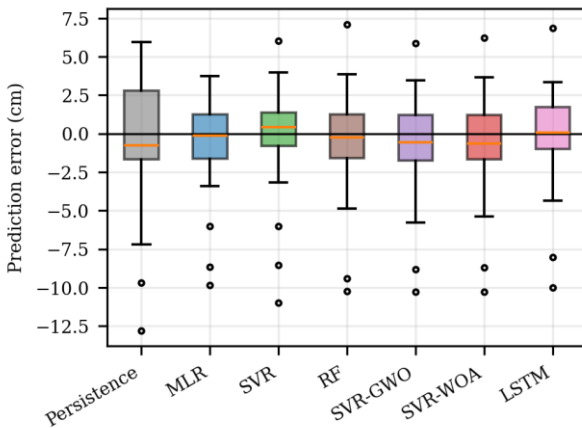


Figure 9. Distribution of prediction errors in the test period

3.3. Importance of Predictors

The permutation importance of the RF model (Fig. 10) indicates that the monthly change in storage was controlled mainly by same-month soil moisture (SM(t)) and air temperature (T(t)), with much smaller contributions from the other variables. Soil moisture integrates antecedent precipitation and evapotranspiration and is therefore a physically meaningful proxy of the near-surface storage component observed by GRACE. Temperature controls snowmelt in

spring and evaporative demand in summer. This agrees with the SHAP analysis of Habibi and Tasouji Hassanpour (2025) and with the input selection of Soltani et al. (2022), who also identified soil moisture and temperature as the key predictors of TWSA in this basin.

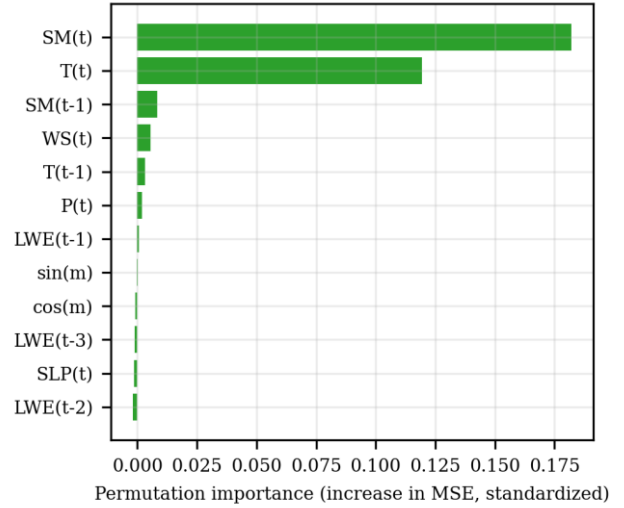


Figure 10. Permutation importance of the predictors (RF, test period)

4. Discussion

The results have three practical implications. First, predicting the monthly increment instead of the storage level was essential. The basin lost more than 20 cm of equivalent water depth during the study period, and the storage in the test period was lower than almost any value seen during training. Learners that cannot extrapolate, such as RF, would have been biased under a level-based formulation. With the increment formulation, all models outperformed persistence, and RF, which had the smallest training error, ranked last among the learners in the test period, a clear sign of overfitting on a small sample.

Second, the advantage of deep learning was small. The LSTM ranked first, but its RMSE was only 1% lower than that of MLR, and it needed about 17 s of training for five ensemble members compared with milliseconds for MLR. With fewer than 120 training samples, the ability of recurrent networks to learn long-term dependencies cannot be fully exploited. Similar conclusions were reached by Cheng et al. (2020), who reported that LSTM loses its advantage at monthly time scales because of the limited size of monthly datasets. The ensemble averaging over five random seeds was important for obtaining a stable LSTM, since individual networks varied noticeably.

Third, the metaheuristic hybrids were a useful and cheap improvement over a default SVR, but they did not outperform the best benchmark models. Both optimizers moved SVR towards a nearly linear model, which explains why their performance approached that of MLR. This does not argue against hybrid modelling, but

it shows that the gain from hyperparameter optimization depends on how far the default configuration is from a suitable one. Improvements reported in the literature (e.g. Seifi et al., 2020; Mozaffari et al., 2022) should therefore be interpreted relative to strong baselines and independent test periods.

The study has several limitations. The analysis was restricted to 2003–2017 because the shared GRACE-FO segment appears inconsistent with the GRACE segment. Extending the models to recent years will require a homogenized TWSA series, for example from the official JPL or CSR mascon releases. The basin-average predictors hide strong spatial contrasts between the mountainous rim and the irrigated plains, and anthropogenic drivers such as groundwater abstraction and reservoir operation, which are known to dominate storage losses in the basin (Joodaki et al., 2014; Sabzehee et al., 2023), were not included because consistent monthly data are not publicly available. Finally, a test period of 30 months is short, and the ranking of models with very similar errors should not be over-interpreted.

5. Conclusions

This study developed and compared MLR, SVR, RF, LSTM and two metaheuristic-optimized SVR models (SVR-GWO and SVR-WOA) for one-month-ahead estimation of GRACE-derived water storage anomalies in the Lake Urmia basin. The GRACE-period record showed a highly significant decline of about 2.0 cm per year. Under a strictly chronological evaluation, all learning models improved on persistence. The LSTM achieved the best test accuracy (RMSE = 3.16 cm, NSE = 0.76), but MLR and the two hybrid SVR models were almost as accurate. GWO and WOA reduced the test error of the default SVR by about 2% at negligible computational cost. Soil moisture and air temperature were the dominant predictors.

For operational use in the Lake Urmia basin, a simple ensemble of a linear model, a tuned SVR and an LSTM trained on storage increments is recommended. Future work should use homogenized GRACE/GRACE-FO mascon solutions, add groundwater abstraction and reservoir data, and quantify forecast uncertainty.

6. Data Availability

The monthly data were obtained from the public repository github.com/habibisara/Urmia-Lake-Climate-Study (Habibi and Tasouji Hassanpour, 2025). Original sources: GRACE (NASA/JPL, via Google Earth Engine), Iran Meteorological Organization and GLDAS (NASA GSFC). 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.

- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Cheng, M., et al. (2020). Long lead-time daily and monthly streamflow forecasting using machine learning methods. *Journal of Hydrology*, 590, 125376.
- Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support vector regression machines. *Advances in Neural Information Processing Systems*, 9, 155–161.
- Foroumandi, E., Nourani, V., Jeanne Huang, J., & Moradkhani, H. (2023). Drought monitoring by downscaling GRACE-derived terrestrial water storage anomalies: A deep learning approach. *Journal of Hydrology*, 616, 128838.
- 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.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Joodaki, G., Wahr, J., & Swenson, S. (2014). Estimating the human contribution to groundwater depletion in the Middle East, from GRACE data, land surface models, and well observations. *Water Resources Research*, 50(3), 2679–2692.
- Kayhomayoon, Z., et al. (2022). A combination of metaheuristic optimization algorithms and machine learning methods improves the prediction of groundwater level. *Water*, 14(5), 751.
- Kendall, M. G. (1975). *Rank Correlation Methods* (4th ed.). Charles Griffin, London.
- Khorrami, B., et al. (2023). Model-coupled GRACE-based analysis of hydrological dynamics of drying Lake Urmia and its basin. *Hydrological Processes*, 37, e14893.
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences*, 22(11), 6005–6022.
- Landerer, F. W., Flechtner, F. M., Save, H., et al. (2020). Extending the global mass change data record: GRACE Follow-On instrument and science data performance. *Geophysical Research Letters*, 47(12), e2020GL088306.
- Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica*, 13(3), 245–259.
- Mirjalili, S., & Lewis, A. (2016). The whale optimization algorithm. *Advances in Engineering Software*, 95, 51–67.
- Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. *Advances in Engineering Software*, 69, 46–61.
- Mozaffari, S., et al. (2022). Forecasting groundwater levels using a hybrid of support vector regression and particle

- swarm optimization. *Water Resources Management*, 36(6). <https://doi.org/10.1007/s11269-022-03118-z>
- 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.
- Radman, A., et al. (2022). Monitoring and predicting temporal changes of Urmia Lake and its basin using satellite multi-sensor data and deep-learning algorithms. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*. <https://doi.org/10.1007/s41064-022-00203-1>
- 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.
- Sabzehee, F., Amiri-Simkooei, A. R., Iran-Pour, S., Vishwakarma, B. D., & Kerachian, R. (2023). Enhancing spatial resolution of GRACE-derived groundwater storage anomalies in Urmia catchment using machine learning downscaling methods. *Journal of Environmental Management*, 330, 117180.
- Seifi, A., Ehteram, M., Singh, V. P., & Mosavi, A. (2020). Modeling and uncertainty analysis of groundwater level using six evolutionary optimization algorithms hybridized with ANFIS, SVM, and ANN. *Sustainability*, 12(10), 4023.
- Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, 63(324), 1379–1389.
- Soltani, K., et al. (2022). Forecasting groundwater anomaly in the future using satellite information and machine learning. *Journal of Hydrology*, 612, 128052.
- Soltani, K., et al. (2024). Terrestrial water storage anomaly estimating using machine learning techniques and satellite-based data (a case study of Lake Urmia Basin). *Irrigation and Drainage*. <https://doi.org/10.1002/ird.2863>
- Sun, A. Y., Scanlon, B. R., Save, H., & Rateb, A. (2020). Reconstruction of GRACE total water storage through automated machine learning. *Water Resources Research*, 56(12), e2020WR028666.
- Tapley, B. D., Bettadpur, S., Watkins, M., & Reigber, C. (2004). The gravity recovery and climate experiment: Mission overview and early results. *Geophysical Research Letters*, 31(9), L09607.
- Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, 106(D7), 7183–7192.
- Tourian, M. J., Elmi, O., Chen, Q., Devaraju, B., Roohi, S., & Sneeuw, N. (2015). A spaceborne multisensor approach to monitor the desiccation of Lake Urmia in Iran. *Remote Sensing of Environment*, 156, 349–360.
- Van Thieu, N., & Mirjalili, S. (2023). MEALPY: An open-source library for latest meta-heuristic algorithms in Python. *Journal of Systems Architecture*, 139, 102871.
- Voss, K. A., Famiglietti, J. S., Lo, M., de Linage, C., Rodell, M., & Swenson, S. C. (2013). Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris-Euphrates-Western Iran region. *Water Resources Research*, 49(2), 904–914.
- Willmott, C. J. (1981). On the validation of models. *Physical Geography*, 2(2), 184–194.
- Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1), 67–82.