

Satellite-Derived Storage Dynamics and Multi-Horizon Forecasting of the Chah-Nimeh 3 Reservoir (Sistan, Iran) with Machine Learning, LSTM and Metaheuristic-Tuned Models

Hadi SIASAR^{1*}✉, Iman Homayunnejad²

¹ Assistant Professor of Department of Agircultural, Payame Noor University, Tehran, Iran. (✉ hadisiasar@pnu.ac.ir)

² Assistant Professor, Department of Agriculture, Faculty of Engineering, Payame Noor University

Received: 06 May 2025

Accepted: 18 June 2025

Available online: 21 June 2025

Abstract: The Chah-Nimeh reservoirs store the water that supplies the Sistan plain in south-eastern Iran, but their storage is not publicly monitored in real time. This study analyses a monthly satellite-derived record of the Chah-Nimeh 3 reservoir (December 2018 – December 2025, 84 months) that contains water-surface area, water-surface elevation, satellite-derived depth and four storage estimates (linear, quadratic, spectral and hybrid area–volume models), and develops one- and three-month-ahead forecasting models for the hybrid storage. Multiple linear regression, support vector regression (SVR), random forest (RF), an LSTM network and two hybrid models tuned by the grey wolf optimizer (SVR-GWO) and Harris hawks optimization (XGB-HHO) were compared with a persistence benchmark under a chronological split (training 2018–2022, validation 2023, test 2024–2025). Storage declined significantly over the record (Mann–Kendall $Z = -4.87$, Sen's slope = -23.3 MCM/year), from an annual mean of 311 MCM in 2020 to 120 MCM in 2023, before recovering to 245 MCM in 2025. The four volume estimators were highly correlated ($r \geq 0.96$) but differed by more than a factor of two in magnitude, the spectral estimator being the highest. At a one-month horizon the LSTM was the most accurate model (RMSE = 34.4 MCM, NSE = 0.72), marginally better than persistence (37.4 MCM). At three months the advantage of learning models became clearer: RF reached RMSE = 62.1 MCM and NSE = 0.07 against 68.4 MCM and -0.12 for persistence. Lagged storage, surface area and the seasonal terms were the dominant predictors. The results show that satellite storage records can support short-term reservoir operation, and that the benefit of data-driven models grows with the forecast horizon.

Keywords: Reservoir storage, Satellite bathymetry, Chah-Nimeh, LSTM, Grey wolf optimizer, Harris hawks optimization

1. Introduction

Reservoirs are the backbone of water supply in arid regions, yet their storage is often not measured or not published. Satellite remote sensing offers an independent alternative: the water-surface area can be mapped with optical imagery, the water level can be obtained from altimetry or from a digital elevation model, and their combination yields storage variations (Gao, 2015). Global implementations of this idea (Duan and Bastiaanssen, 2013; Zhang et al., 2014; Busker et al., 2019; Tortini et al., 2020) have shown that satellite storage estimates can approach the accuracy of in-situ records for large reservoirs.

For small and shallow water bodies, an additional source of information is satellite-derived bathymetry. Log-ratio approaches relate the ratio of logarithms of two visible bands to depth (Lyzenga, 1978; Stumpf et al., 2003) and can be inverted to obtain depth maps and therefore volumes. Combining an area–elevation relationship with a spectral depth estimate produces several possible volume estimators, and their differences are rarely quantified.

Once a storage record exists, the natural next step is forecasting. Machine learning models have been widely applied to reservoir inflow and storage prediction, and recurrent networks have become the reference method for monthly water-resources series (Hochreiter and Schmidhuber, 1997; Kratzert et al., 2018). Metaheuristic algorithms such as the grey wolf optimizer (Mirjalili et al., 2014) and Harris hawks optimization (Heidari et al.,

2019) are often used to tune these models, and reported improvements are large. Most studies, however, evaluate only one forecast horizon and compare their models with untuned alternatives rather than with a persistence benchmark, which is a strong competitor for a slowly varying quantity such as reservoir storage.

The Chah-Nimeh reservoirs supply drinking water to Zabol and to a large agricultural area of the Sistan plain, and their filling depends on the transboundary Hirmand (Helmand) River. The objectives of this study were (i) to describe the storage dynamics of Chah-Nimeh 3 from a satellite-derived monthly record and to compare four area–volume estimators, (ii) to forecast the reservoir storage one and three months ahead with MLR, SVR, RF, LSTM, SVR-GWO and XGB-HHO, and (iii) to quantify the gain over a persistence benchmark as a function of the forecast horizon.

2. Materials and Methods

2.1. Study Area and Data

Chah-Nimeh 3 (about 30.83° N, 61.69° E) is one of four natural depressions east of the Sistan plain that are filled by diversion from the Hirmand River (Fig. 1). The climate is hyper-arid, with less than 60 mm of annual precipitation and a potential evapotranspiration above 3,000 mm, so that evaporation losses from the reservoir surface are substantial. The reservoirs are the last large water store of a basin whose terminal wetlands, the Hamoun lakes, have repeatedly dried out during the past two decades (Sharifikia, 2013).

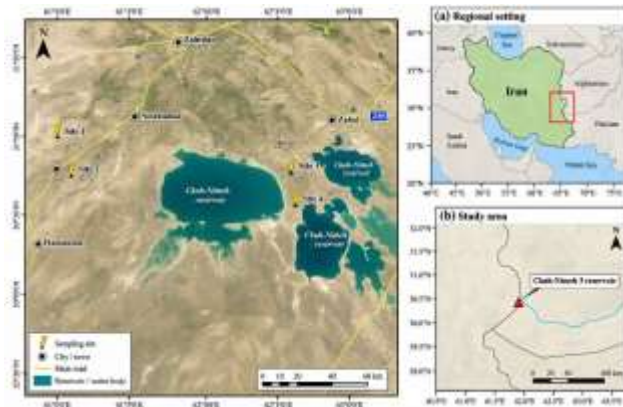


Figure 1. Location of the Chah-Nimeh 3 reservoir, Sistan, Iran

The dataset used here consists of 84 consecutive monthly records (December 2018 – December 2025) derived from Sentinel-2 imagery processed in Google Earth Engine (Gorelick et al., 2017). Each record contains the water-surface area obtained with the NDWI, the water-surface elevation and the corresponding water depth, a spectrally derived mean depth based on the log-ratio algorithm of Stumpf et al. (2003), and four storage estimates: a linear and a quadratic area–volume interpolation, a purely spectral estimate (mean spectral depth multiplied by area) and a hybrid estimate that combines the geometric and spectral information. The hybrid estimate is used as the target variable, following the recommendation of the data producers. Descriptive statistics are given in Table 1.

Table 1. Descriptive statistics of the monthly satellite record (December 2018 – December 2025)

Variable	Min	Max	Mean	SD	Skew	Kurt
Hybrid volume (MCM)	12.77	375.10	228.19	82.63	-0.38	-0.40
Quadratic (MCM)	9.40	307.52	168.62	76.62	-0.26	-0.61
Linear (MCM)	1.47	313.36	166.46	77.78	0.06	-0.74
Spectral (MCM)	39.13	519.14	372.23	100.95	-1.04	0.76
Area (km ²)	1.37	18.32	13.48	3.45	-1.12	1.08
Elevation (m)	461.52	491.12	482.09	6.18	-0.99	0.72

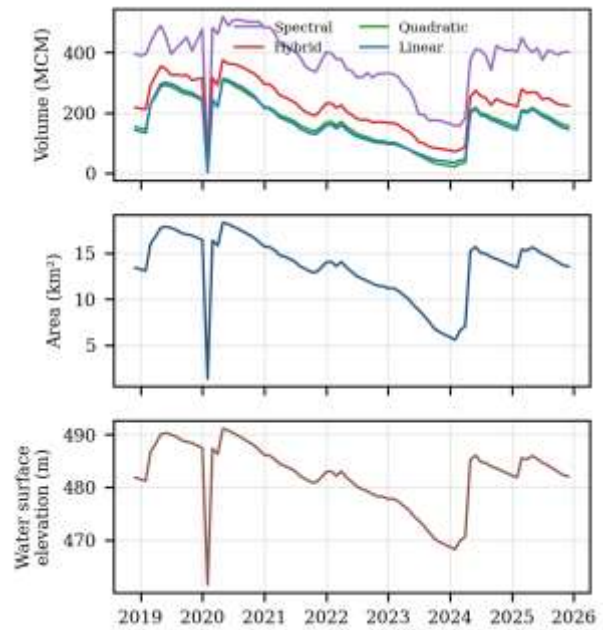


Figure 2. Monthly storage from the four estimators, water-surface area and water-surface elevation

2.2. Forecasting Setup

The target was the hybrid storage of month $t+h$ with $h = 1$ and $h = 3$ months. Because storage is strongly autocorrelated and non-stationary, all models were trained to predict the storage increment $V(t+h) - V(t)$, and the forecast was obtained by adding the predicted increment to the last observed storage. The predictors were the storage of months t , $t-1$ and $t-2$, its one- and three-month changes, the water-surface area and its change, the water-surface elevation and its change, the spectral mean depth and the sine and cosine of the target month. The LSTM received the previous 12 months of $[V, \text{area}, \text{elevation}, \text{spectral depth}, \sin, \cos]$.

The record was split chronologically by target month into training (to December 2022, $n = 44$), validation (2023, $n = 12$) and test (2024–2025, $n = 24$) sets; the three-month models used 42, 12 and 24 samples. All inputs were standardized with training statistics, and a persistence forecast ($V(t+h) = V(t)$) served as benchmark.

2.3. Models

MLR, **SVR** with a radial basis function kernel (Drucker et al., 1997) and **RF** with 500 trees (Breiman, 2001) were used as baselines. The **LSTM** had one layer of 32 units, dropout of 0.2 and a 16-unit dense layer, was trained with Adam and early stopping, and five networks with different seeds were averaged. In the hybrid models, the SVR hyperparameters (C , γ , ϵ) were tuned by the **grey wolf optimizer** and the XGBoost hyperparameters by **Harris hawks optimization**, with 20 agents and 40 iterations and the validation RMSE as fitness (MEALPY; Van Thieu and Mirjalili, 2023). GWO updates each search agent towards the three best solutions, while HHO alternates between exploration

and besiege phases controlled by the escaping energy $E = 2E_0(1 - t/T)$.

2.4. Evaluation

Forecasts were evaluated in the original units with R^2 , RMSE, MAE, NSE (Nash and Sutcliffe, 1970), KGE (Gupta et al., 2009) and Willmott's index (Willmott, 1981):

$$NSE = 1 - \Sigma(V_o - V_s)^2 / \Sigma(V_o - \bar{V}_o)^2 \quad (1)$$

$$KGE = 1 - \sqrt{[(r-1)^2 + (\sigma_s/\sigma_o - 1)^2 + (\mu_s/\mu_o - 1)^2]} \quad (2)$$

3. Results

3.1. Storage Dynamics

The reservoir passed through a complete depletion–recovery cycle during the study period (Fig. 2). The hybrid storage reached 375 MCM in May 2020, when the water surface covered 18.3 km² at an elevation of 491.1 m, and fell to a minimum of 12.8 MCM in February 2020. The annual means decreased from 311 MCM in 2020 to 120 MCM in 2023 (–61%) and recovered to 190 and 245 MCM in 2024 and 2025 after the spring floods of 2024. The Mann–Kendall test indicates a significant declining trend over the whole record ($Z = -4.87$, $p < 0.001$, Sen's slope = -23.3 MCM/year), which mainly reflects the difference between the wet beginning and the dry middle of the period. The extremely low value of February 2020, which coincides with the lowest recorded area (1.37 km²) three months before the maximum, is not consistent with the neighbouring months and should be regarded as an artefact of the monthly composite rather than a real drawdown.

The four storage estimators are highly correlated (Table 2): the hybrid estimate correlates at $r = 0.993$ with the quadratic, 0.986 with the linear and 0.955 with the spectral estimate. Their magnitudes, however, differ strongly: the mean storage is 166 MCM (linear), 169 MCM (quadratic), 228 MCM (hybrid) and 372 MCM (spectral). The spectral estimator is about 2.2 times the quadratic one and reacts less to the drawdown, because the mean spectral depth varies only between 22.1 and 30.9 m. Choosing an estimator therefore changes the absolute storage much more than its relative dynamics, and absolute values should be calibrated against in-situ measurements before they are used in a water balance.

Table 2. Correlation of the storage estimators with the hybrid estimate, the area and the elevation, and their mean and standard deviation (MCM)

Estimator	r (hybrid)	r (area)	r (elev.)	Mean	SD
Lin	0.986	0.949	0.958	166.5	77.8
Quad	0.993	0.971	0.978	168.6	76.6
Spec	0.955	0.959	0.958	372.2	100.9
Hyb	1.000	0.977	0.982	228.2	82.6

Storage is a smooth function of both area and elevation (Fig. 3). A quadratic fit of the hybrid storage

on the area explains the data closely, as does a quadratic fit on the elevation, and the area and elevation are themselves almost perfectly related ($r = 0.999$). The seasonal climatology (Fig. 4) shows the highest mean storage in late spring and early summer, after the arrival of the Hirmand floods, and the lowest storage in late winter, which reflects the combination of the inflow regime and the very high evaporation of the summer months.

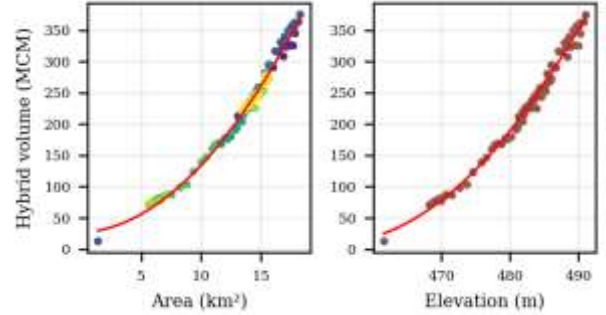


Figure 3. Hybrid storage as a function of (a) water-surface area and (b) water-surface elevation (red: quadratic fit)

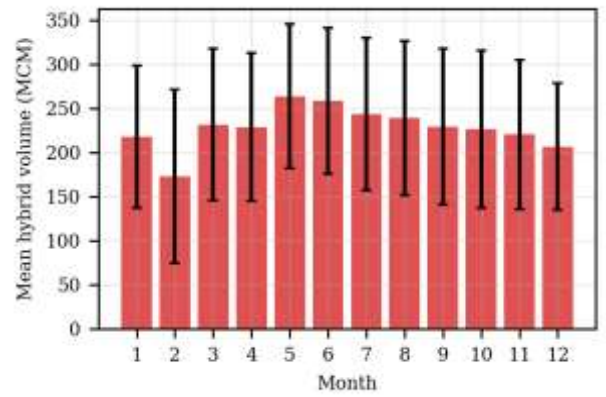


Figure 4. Monthly climatology of the hybrid storage (bars: mean $\pm$ 1 SD)

3.2. Forecast Accuracy

Table 3. Test-period (2024–2025) performance for one- and three-month-ahead storage forecasts; RMSE in MCM

Model	R^2 (h=1)	RMSE (h=1)	NSE (h=1)	KGE (h=1)	RMSE (h=3)	NSE (h=3)
Persistence	0.727	37.4	0.663	0.826	68.4	-0.125
MLR	0.060	72.9	-0.279	0.208	78.5	-0.484
SVR	0.710	36.4	0.682	0.842	65.1	-0.020
RF	0.723	34.8	0.709	0.798	62.1	0.073
SVR-GWO	0.727	39.3	0.629	0.820	75.6	-0.377
XGB-HHO	0.734	36.8	0.674	0.845	71.7	-0.239
LSTM	0.732	34.4	0.715	0.849	62.4	0.061

At the one-month horizon (Table 3), persistence is already a strong benchmark (RMSE = 37.4 MCM, NSE = 0.66), because the storage changes slowly. The LSTM gave the lowest error (RMSE = 34.4 MCM, NSE =

0.715, KGE = 0.849), followed by RF (34.8 MCM), SVR (36.4 MCM) and XGB-HHO (36.8 MCM); the improvement of the best model over persistence is 8%. SVR-GWO (39.3 MCM) did not improve on the untuned SVR in the test period, although it had the lower validation fitness, and MLR failed (NSE = -0.28) because it extrapolated a linear trend of the increments.

At the three-month horizon the errors of all models increase, but the ranking changes in favour of the learning models. RF becomes the best model (RMSE = 62.1 MCM, NSE = 0.073), closely followed by the LSTM (62.4 MCM) and SVR (65.1 MCM), whereas persistence drops to RMSE = 68.4 MCM and a negative efficiency (-0.12). The gain of RF over persistence is 9%. The tuned models again did not outperform their baselines at this horizon (XGB-HHO: 71.7 MCM; SVR-GWO: 75.6 MCM), which suggests that with about 40 training samples the validation set is too small to select hyperparameters reliably.

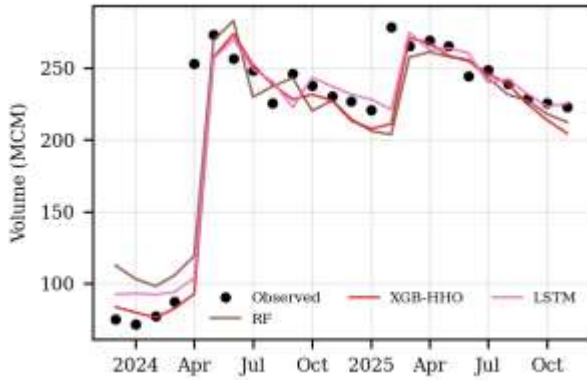


Figure 5. Observed and one-month-ahead forecast storage in the test period

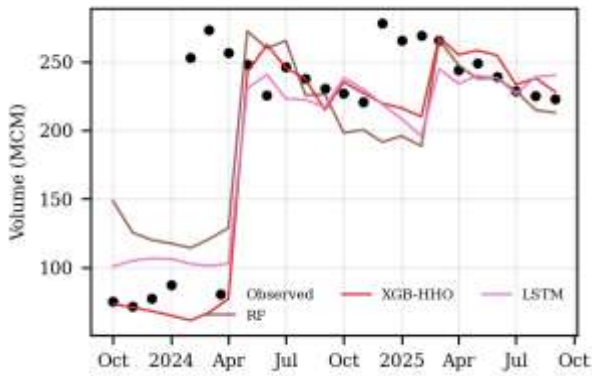


Figure 6. Observed and three-month-ahead forecast storage in the test period

The forecast hydrographs (Figs. 5 and 6) show that the models follow the refilling of 2024 and the subsequent decline, but that they systematically smooth the sharp rise of spring 2024. This is expected: the refilling is driven by releases from upstream and by floods of the Hirmand River, which are not part of the input set. The scatter plots (Fig. 7), the Taylor diagram (Fig. 8) and the error distributions (Fig. 11) confirm that

the differences among the leading models are small compared with their common bias during rapid transitions.

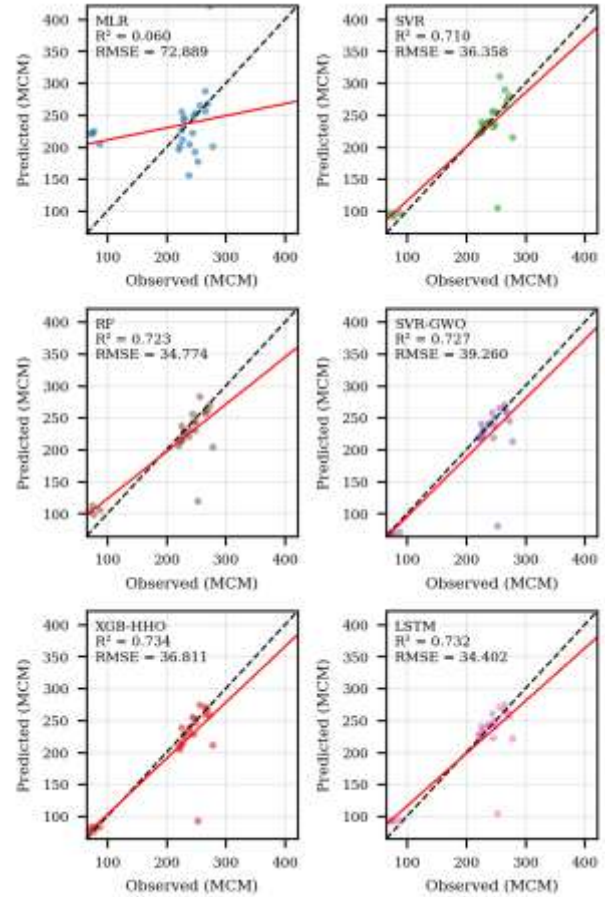


Figure 7. Observed versus one-month-ahead forecast storage in the test period

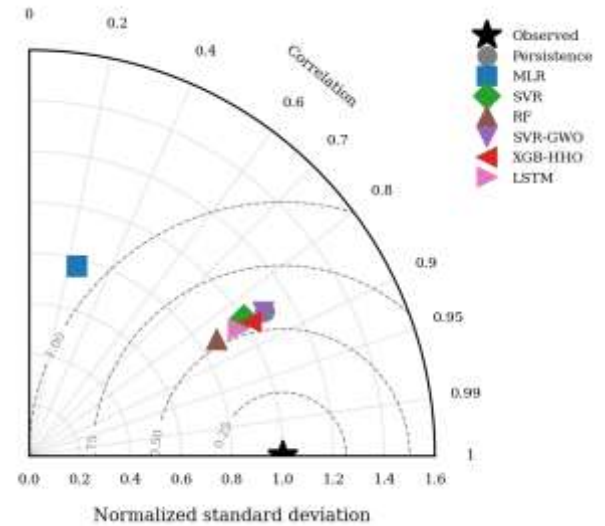


Figure 8. Taylor diagram of the one-month-ahead forecasts

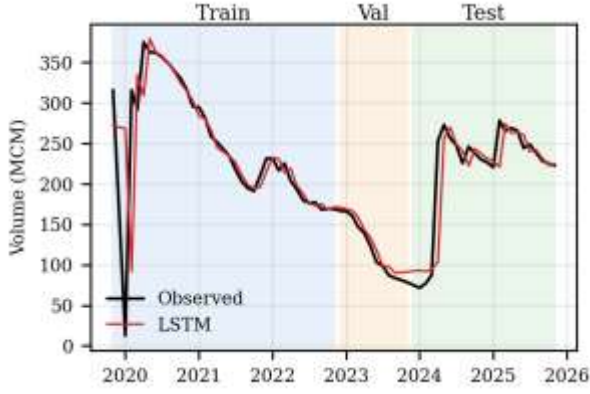


Figure 9. Observed and LSTM-forecast storage over the full record ($h = 1$)

The metaheuristic searches converged quickly (Fig. 10), reducing the standardized validation RMSE from 0.0890 to 0.0888 for SVR-GWO and from 0.1216 to 0.1136 for XGB-HHO, at a cost of 0 s and 12 s. The permutation importance of the random forest (Fig. 11) ranks the lagged storage values ($V(t-2)$, $V(t)$) first, followed by the seasonal term and the change of the water-surface area, confirming that the recent state of the reservoir and the season carry most of the predictive information.

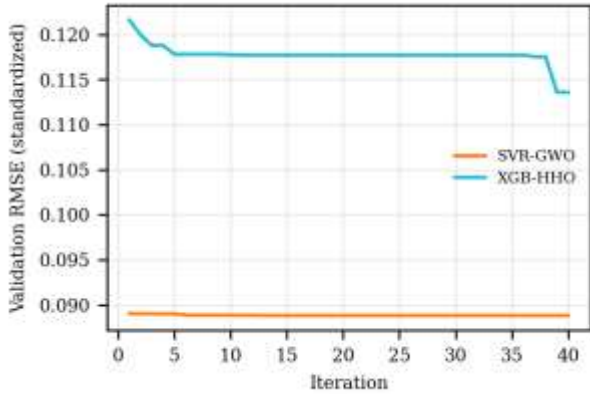


Figure 10. Convergence of the GWO and HHO searches ($h = 1$)

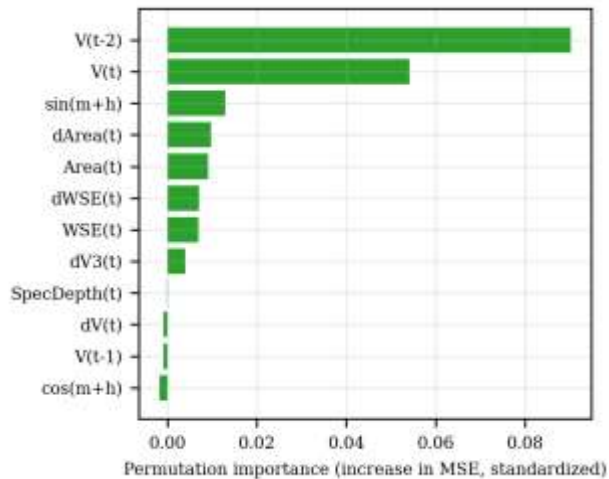


Figure 11. Permutation importance of the random forest inputs ($h = 1$)

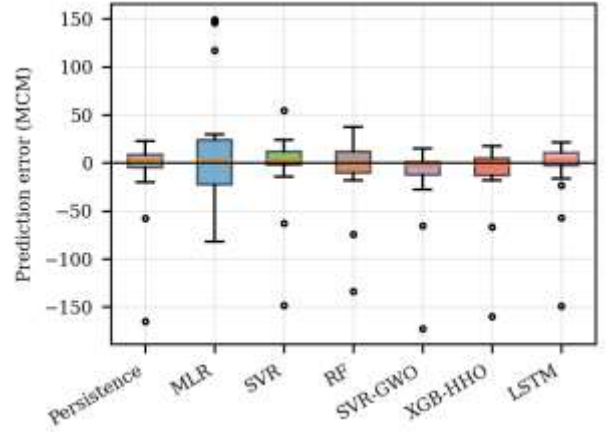


Figure 12. Distribution of one-month-ahead forecast errors in the test period

4. Discussion

The satellite record shows that Chah-Nimeh 3 lost about two thirds of its storage between 2020 and 2023 and recovered a large part of it in 2024. Such a cycle is typical of a reservoir fed by a highly variable transboundary river in a hyper-arid climate, and it demonstrates the value of an independent satellite record for a system whose in-situ data are not published. The comparison of the four estimators shows, however, that the absolute storage remains uncertain: estimators that are mutually correlated at $r > 0.95$ differ by a factor of more than two in magnitude. Users should therefore treat the satellite storage as a relative indicator unless it has been anchored on measured levels or a bathymetric survey.

The forecasting experiments illustrate a general property of reservoir storage: at short horizons the system is dominated by its own memory, so persistence is hard to beat, and the added value of machine learning grows with the lead time. The one-month gain of the best model was modest (about 8%), while at three months the random forest improved on persistence by about 9% in RMSE and turned a negative efficiency into a positive one. For operational use, a three-month forecast is also more useful, because it covers the typical planning horizon for irrigation releases.

Metaheuristic tuning did not help in this application. Both optimizers reduced the validation error but produced slightly worse test forecasts, which is a consequence of the small sample: with 44 training and 12 validation months, the selection of seven hyperparameters is poorly constrained. This does not contradict studies that report large gains for longer records, but it shows that hybrid models should not be applied uncritically to short satellite records.

The main limitation is the absence of inflow information. Including releases from upstream dams, Hirmand discharge or precipitation over the Helmand basin would very likely improve the forecasts of the

refilling episodes, which are the most difficult and the most important events to predict. The record is also short (84 months) and covers only one depletion–recovery cycle, so the models have seen very few analogues of the events they must forecast.

5. Conclusions

A monthly satellite record of the Chah-Nimeh 3 reservoir was analysed and used to forecast storage one and three months ahead. The reservoir lost about 61% of its storage between 2020 and 2023 and recovered to 245 MCM in 2025; the trend over the whole record is significantly negative (-23.3 MCM/year). The four area–volume estimators were strongly correlated but differed by more than a factor of two in magnitude, so satellite storage should be calibrated before absolute use. At a one-month horizon the LSTM was the most accurate model (NSE = 0.72), only 8% better than persistence, while at three months the random forest clearly outperformed persistence (NSE = 0.07 against -0.12). Metaheuristic tuning did not improve the test accuracy for this short record. Future work should incorporate upstream inflow and release information and validate the satellite storage against in-situ levels.

6. Data Availability

The monthly satellite-derived area, elevation, depth and storage series were produced in Google Earth Engine from Sentinel-2 imagery and are available from the authors on request together with the analysis code.

7. References

- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Busker, T., de Roo, A., Gelati, E., Schwatke, C., Adamovic, M., Bisselink, B., Pekel, J.-F., & Cottam, A. (2019). A global lake and reservoir volume analysis using a surface water dataset and satellite altimetry. *Hydrology and Earth System Sciences*, 23(2), 669–690.
- 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.
- Duan, Z., & Bastiaanssen, W. G. M. (2013). Estimating water volume variations in lakes and reservoirs from four operational satellite altimetry databases and satellite imagery data. *Remote Sensing of Environment*, 134, 403–416.
- Gao, H. (2015). Satellite remote sensing of large lakes and reservoirs: From elevation and area to storage. *WIREs Water*, 2(2), 147–157.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27.
- Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1–2), 80–91.
- 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.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- 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.
- Lyzenga, D. R. (1978). Passive remote sensing techniques for mapping water depth and bottom features. *Applied Optics*, 17(3), 379–383.
- Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. *Advances in Engineering Software*, 69, 46–61.
- 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.
- Sharifikia, M. (2013). Environmental challenges and drought hazard assessment of Hamoun Desert Lake in Sistan region, Iran, based on the time series of satellite imagery. *Natural Hazards*, 65, 201–217.
- Stumpf, R. P., Holderied, K., & Sinclair, M. (2003). Determination of water depth with high-resolution satellite imagery over variable bottom types. *Limnology and Oceanography*, 48(1, part 2), 547–556.
- Tortini, R., Noujdina, N., Yeo, S., Ricko, M., Birkett, C. M., Khandelwal, A., Kumar, V., Marlier, M. E., & Lettenmaier, D. P. (2020). Satellite-based remote sensing data set of global surface water storage change from 1992 to 2018. *Earth System Science Data*, 12(2), 1141–1151.
- Van Thieu, N., & Mirjalili, S. (2023). MEALPY: An open-source library for latest meta-heuristic algorithms in Python. *Journal of Systems Architecture*, 139, 102871.
- Willmott, C. J. (1981). On the validation of models. *Physical Geography*, 2(2), 184–194.
- Zhang, S., Gao, H., & Naz, B. S. (2014). Monitoring reservoir storage in South Asia from multisatellite remote sensing. *Water Resources Research*, 50(11), 8927–8943.