International Journal of Environment, Agriculture and Biotechnology
Vol-11, Issue-4; Jul-Aug, 2026
Peer-Reviewed International Journal
Journal Home Page Available: https://ijeab.com/
Journal DOI: 10.22161/ijeab

Prediction of Daily Water Level Variations in Thurmond Reservoir Using Classical and Artificial Intelligence Methods

Aysu Ayar*, Fatih Üneş, Bestami Taşar

Department of Civil Engineering, Iskenderun Technical University, Türkiye
*Email: aysuayar123@gmail.com

Received: 18 Jun 2026; Received in revised form: 12 Jul 2026; Accepted: 16 Jul 2026; Available online: 22 Jul 2026

***Abstract*— *Accurate prediction of reservoir water levels is essential for sustainable water resources management, reservoir operation, and water resources planning under increasing water demand and climate variability. In this study, the daily water level of Thurmond Reservoir, located on the Savannah River between South Carolina and Georgia, USA, was predicted using daily average air temperature (T), relative humidity (RH), precipitation (P), and one-day lagged reservoir water level [RWL(t−1)]. A total of 1608 daily observations collected between 2017 and 2023 were used, with 80% of the dataset allocated for training and the remaining 20% for testing. Multiple Linear Regression (MLR), Interaction Regression (IR), Quadratic Regression (QR), and Long Short-Term Memory (LSTM) models were employed to estimate daily reservoir water levels. Model performances were evaluated using the Coefficient of Determination (R^2), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). The results demonstrated that all developed models achieved satisfactory performance for daily reservoir water-level prediction. Among the evaluated models, the QR model achieved the highest prediction accuracy and the lowest error. The results obtained show that the proposed models can be effectively used for reservoir operation, water resources planning, and sustainable reservoir management.*

Keywords— **Reservoir water level, daily water level prediction, multiple linear regression, LSTM, water resources management.**

I. INTRODUCTION

Freshwater resources are increasingly threatened by climate change, population growth, urbanization, and rising water demand. These challenges have highlighted the necessity for effective planning and sustainable management of available water resources [1]. Changes in precipitation patterns, the increasing frequency and severity of droughts, and growing water demand have further underscored the importance of efficient reservoir operation and water resources management. Therefore, accurate prediction of reservoir water levels has become essential for reservoir operation, water resources planning, hydropower generation, flood control, and sustainable water management.

Reservoir water levels are affected by numerous hydrological and meteorological variables, including precipitation, air temperature, relative humidity, evaporation, streamflow, and reservoir operation policies. The complex and nonlinear interactions among these variables make daily reservoir water level prediction a challenging task in hydrological studies. Reliable prediction of reservoir water-level fluctuations contributes significantly to reservoir operations, drought and flood risk mitigation, and the efficient utilization of water resources.

Traditionally, reservoir water level prediction has been performed using statistical approaches, regression models, and physically based hydrological models. However, conventional methods may fail to accurately represent the nonlinear characteristics of hydrological processes [2]. In recent years, Artificial Intelligence techniques have gained considerable attention because of their ability to capture complex nonlinear relationships and improve prediction accuracy. Nevertheless, classical regression models remain attractive owing to their simple mathematical structure, ease of implementation, and interpretability.

In recent decades, artificial intelligence and machine learning methods have been widely applied to various hydrological and water-resources problems, including reservoir volume estimation, reservoir and lake water-level prediction, rainfall–runoff modeling, discharge forecasting, and reference evapotranspiration estimation. Generalized regression neural networks, support vector machines, M5 decision trees, adaptive neuro-fuzzy inference systems, artificial neural networks, deep neural networks, long short-term memory networks, and ensemble learning approaches have shown strong potential for representing complex and nonlinear hydrological processes [3-15]. These studies indicate that artificial intelligence-based models provide reliable alternatives for accurate water-level prediction. Özdemir et al. [16] compared Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Stacked LSTM, and Bidirectional LSTM models for predicting the water level of Sapanca Lake and reported that the LSTM model provided higher prediction accuracy, particularly for long-term forecasting. Zhu et al. [17] evaluated Feedforward Neural Networks (FFNNs), GRUs, LSTMs, and attention-based deep learning models for daily reservoir water-level prediction in five lakes in Poland. Their findings demonstrated that all deep learning models achieved satisfactory predictive performance, with LSTM-based models generally achieving the highest accuracy. These studies demonstrate the effectiveness of deep learning techniques for reservoir water level prediction. However, comparative studies that evaluate both classical regression models and deep learning approaches on the same dataset remain limited.

In this study, the daily reservoir water level of Thurmond Reservoir, located on the Savannah River between South Carolina and Georgia, USA, was predicted using daily average air temperature (T), relative humidity (RH), precipitation (P), and one-day lagged reservoir water level [RWL(t−1)] as input variables. Multiple Linear Regression (MLR), Interaction Regression (IR), Quadratic Regression (QR), and Long Short-Term Memory (LSTM) models were developed and compared. Model performance was evaluated using the Coefficient of Determination (R^2), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) to identify the most suitable approach for daily reservoir water-level prediction.

II. STUDY AREA

The present study was conducted using daily reservoir water-level data collected at Thurmond Reservoir, located on the Savannah River along the border between South Carolina and Georgia, USA (approximately 33°39′ N, 82°13′ W). The reservoir was constructed by the U.S. Army Corps of Engineers for multiple purposes, including hydropower generation, flood control, water supply, irrigation, and recreation. Construction of the dam began in 1946 and was completed in 1954. Originally known as Clarks Hill Lake, the reservoir was later renamed J. Strom Thurmond Reservoir. The study area is characterized by a humid subtropical climate with hot, humid summers and mild winters. Seasonal variations in precipitation and temperature, together with reservoir operation policies, considerably influence daily fluctuations in reservoir water levels.

Daily reservoir water level data were obtained from the United States Geological Survey (USGS) database. A total of 1,608 daily observations collected between 2017 and 2023 were used to develop the model. The dataset was randomly divided into 1,286 observations (80%) for training and 322 observations (20%) for testing. Daily average air temperature (T), relative humidity (RH), precipitation (P), and the one-day lagged reservoir water level [RWL(t−1)] were selected as input variables, whereas the daily reservoir water level [RWL(t)] was used as the output variable.

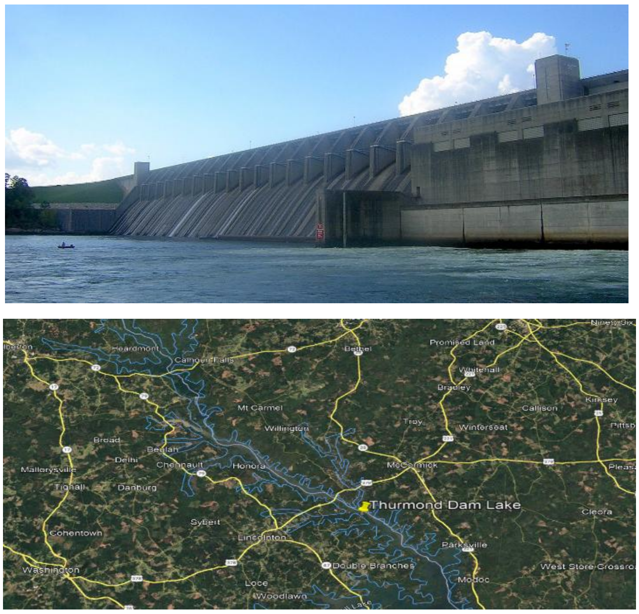

Fig. 1: General view of Thurmond Reservoir [18] and its location [19].

Daily reservoir water level prediction was performed using Multiple Linear Regression (MLR), Interaction Regression (IR), Quadratic Regression (QR), and the Long Short-Term Memory (LSTM) model. The performance of the developed models was evaluated using the Coefficient of Determination (R^2), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). The location, general view and measurement values of the Thurmond Reservoir are presented in Figures 1 and 2, respectively.

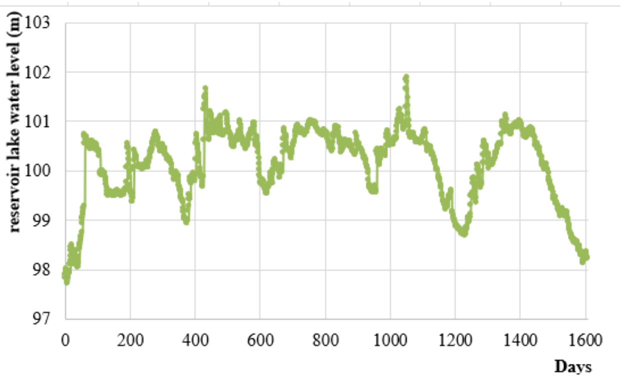

Fig. 2 : Daily reservoir water level data used in this study.

III. METHODS

3.1 Multiple Linear Regression (MLR)

Multiple Linear Regression (MLR) is one of the most widely used statistical techniques for modeling the linear relationship between a dependent variable and multiple independent variables. In this study, the MLR model was employed to predict the daily reservoir water level. The relationship between the dependent and independent variables is expressed by a linear regression equation. The general mathematical form of the MLR model is given in Eq. (1).

$$Z = A_0 + A_1 * X_1 + A_2 * X_2 \dots . + A_i * X_i + \varepsilon \qquad (1)$$

where Z is the dependent variable, X_i represents the independent variables, A_0 is the intercept, A_i denotes the regression coefficients, and ε is the error term.

3.2 Interaction Regression (IR)

Interaction Regression (IR) is a regression technique that considers not only the linear relationship between the dependent and independent variables but also the interaction effects among the independent variables. In addition to the linear terms, interaction terms are included in the regression model to better capture the combined effects of the input variables on the dependent variable. The general mathematical form of the IR model is given in Eq. (2).

$$Z_i = (A_0 + A_1X_1 + A_2X_2 + \cdots + A_nX_n + A_{n+1} * X_1 * X_2 + \cdots) + \varepsilon_i \qquad (2)$$

where Z_i is the dependent variable, X represents the independent variables, A denotes the regression coefficients, and ε_i is the error term [15].

3.3 Quadratic Regression (QR)

Quadratic Regression (QR) is a regression technique used to model both linear and nonlinear relationships between the dependent and independent variables. In addition to the linear terms, the model incorporates interaction and quadratic terms to better represent the individual and combined effects of the input variables on the dependent variable. The general mathematical form of the QR model is given in Eq. (3).

$$Z_i = (A_0 + A_1X_1 + A_2X_2 + \cdots + A_nX_n + A_{n+1}X_1^2 + A_{n+2}X_2^2 + \cdots + A_mX_n^2) + \varepsilon_i \qquad (3)$$

where Z_i is the dependent variable, X represents the independent variables, A denotes the regression coefficients, and ε_i is the error term

3.4 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is an advanced type of Recurrent Neural Network (RNN) designed to overcome the vanishing gradient problem in conventional RNNs and effectively capture long-term temporal dependencies in sequential data. Owing to its memory structure, LSTM has been widely employed in hydrological and meteorological time-series forecasting. The LSTM architecture, originally proposed by Hochreiter and Schmidhuber [20] and further improved by Gers et al. [21], consists of memory cells controlled by three gate mechanisms: the forget gate, input gate, and output gate. These gates regulate the flow of information through the network, allowing relevant historical information to be retained while discarding irrelevant information.

In the present study, the LSTM model was employed to predict the daily reservoir water level of Thurmond Reservoir. Daily average air temperature (T), relative humidity (RH), precipitation (P), and the one-day-lagged reservoir water level [RWL(t−1)] were used as input variables, whereas the daily reservoir water level [RWL(t)] was considered the output variable. The general architecture of the LSTM cell is illustrated in Fig. 3.

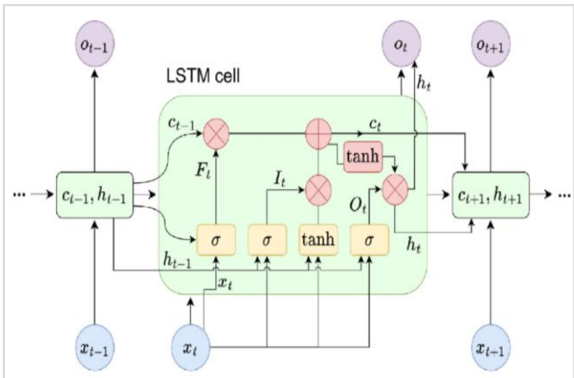

Fig. 3: General architecture of the LSTM cell [22].

The internal operations of the LSTM cell are mathematically expressed in Eqs. (4)–(9), corresponding to the forget gate, input gate, candidate cell state, updated cell state, output gate, and hidden state, respectively.

Ft=σ(WFxt+UFht−1+bF) (4)

$$I_t = \sigma(W_I x_t + U_I h_{t-1} + b_I) \quad (5)$$

$$\tilde{C}_t = \tanh(W_C x_t + U_C h_{t-1} + b_C) \quad (6)$$

$$C_t = F_t \odot C_{t-1} + I_t \odot \tilde{C}_t \quad (7)$$

$$O_t = \sigma(W_O x_t + U_O h_{t-1} + b_O) \quad (8)$$

$$h_t = O_t \odot \tanh(C_t) \quad (9)$$

Where *Ft*, *It,* and *Ot* denote the forget, input, and output gates, respectively; *Ct* represents the cell state, and ht denotes the hidden state. Furthermore, is the input vector at the current time step, while h_{t-1} and C_{t-1}represent the hidden state and cell state from the previous time step. W and U denote the weight matrices, b is the bias term, σ represents the sigmoid activation function, tanh denotes the hyperbolic tangent activation function, and ⊙ indicates element-wise multiplication.

IV. ANALYSIS RESULTS AND DISCUSSION

A total of 1608 daily observations collected between 2017 and 2023 were used in this study. The dataset was randomly divided into 1286 observations (80%) for training and 322 observations (20%) for testing. The performance of the Multiple Linear Regression (MLR), Interaction Regression (IR), Quadratic Regression (QR), and Long Short-Term Memory (LSTM) models was evaluated on the test dataset. Model performance was assessed using the Coefficient of Determination (R^2), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). Lower MAE and RMSE values indicate better model. MAE and RMSE were calculated using Eqs. (10) and (11), respectively. The statistical performance of the developed models is summarized in Table 1.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|RWLi_{measure} - RWLi_{prediction}\right| \quad (10)$$

$$\text{RMSE} = \left(\frac{1}{N}\sum_{i=1}^{N}\left(RWLi_{measure} - RWLi_{prediction}\right)\right)^{1/2} \quad (11)$$

In these equations, N denotes the total number of observations; i denotes the index of each observation; RWLi, measure is the measured reservoir water level for the i-th observation; and RWLi, prediction is the corresponding predicted reservoir water level.

Table 1: Performance comparison of the developed models.

Model	Model Inputs	MAE (m)	RMSE (m)	R²
LSTM	RH(t), T(t), P(t), RWL(t−1)	0.0326	0.0460	0.9973
MLR	RH(t), T(t), P(t), RWL(t−1)	0.0320	0.0460	0.9970
IR	RH(t), T(t), P(t), RWL(t−1)	0.0321	0.0460	0.9973
QR	RH(t), T(t), P(t), RWL(t−1)	0.0313	0.0450	0.9974

The results of the model developed in Table 1 are arranged according to the determined criteria. In order to estimate the daily reservoir water level [RWL(t)], daily mean relative humidity (RH), air temperature (T), precipitation (P) and one-day delayed reservoir water level [RWL(t−1)] were selected as input variables and used.

4.1 MLR Results

In the MLR model, daily average relative humidity (RH), air temperature (T), precipitation (P), and the one-day lagged reservoir water level [RWL(t−1)] were used to predict the daily reservoir water level [RWL(t)]. The time-series comparison between the observed and predicted reservoir water levels for the testing dataset is presented in Fig. 4, while the corresponding scatter plot is shown in Fig. 5.

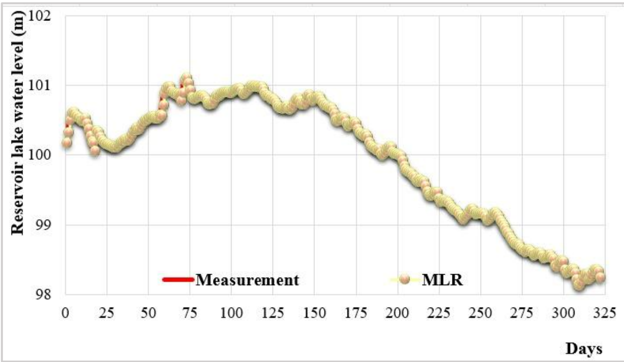

Fig. 4: Comparison of the observed and predicted daily reservoir water levels obtained by the MLR model

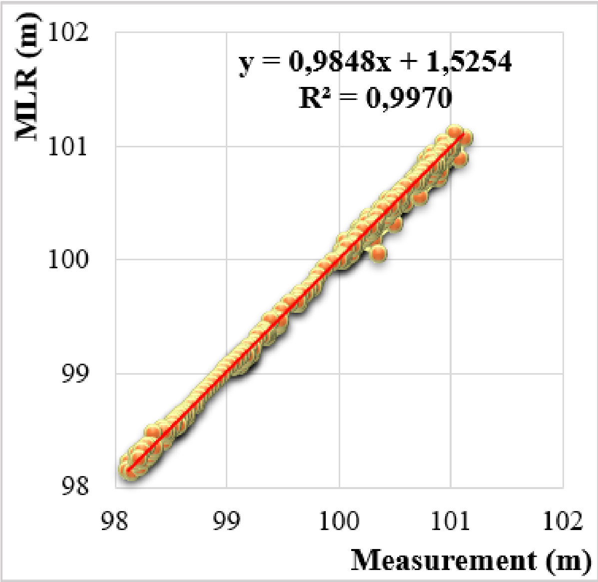

Fig. 5: Scatter plot of the observed versus predicted daily reservoir water levels for the MLR model.

According to the scatter plot (Fig. 5) and Table 1, the MLR model achieved an MAE of 0.0320 m, an RMSE of 0.0460 m, and an R^2 value of 0.9970. These results indicate a high agreement between the observed and predicted reservoir water levels. As shown in Fig. 4, the predicted reservoir water levels closely followed the observed values throughout the testing period. Although minor deviations were observed during periods of relatively rapid water-level fluctuations, the MLR model successfully captured the overall trend in daily reservoir water-level variations.

4.2 (IR) Results

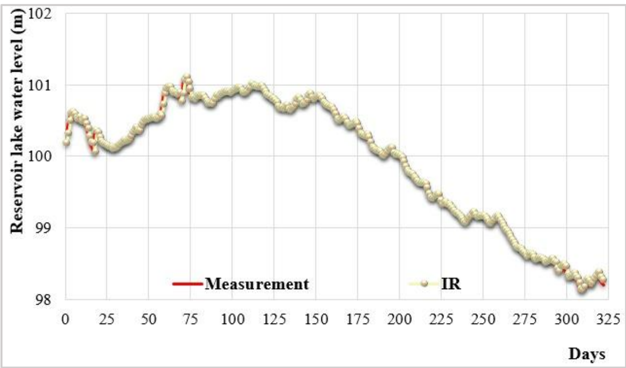

Fig. 6: Comparison between the observed and predicted daily reservoir water levels from the Interaction Regression (IR) model.

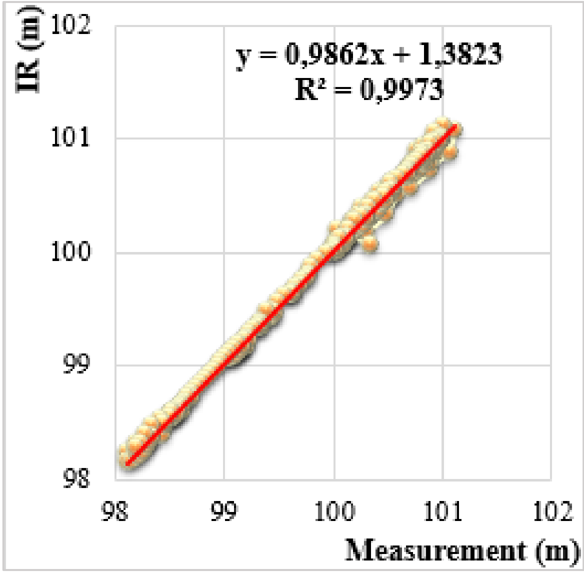

Fig. 7: Scatter plot of the observed versus predicted daily reservoir water levels for the IR model.

According to Table 1, the IR model achieved an MAE of 0.0321 m, an RMSE of 0.0460 m, and an R^2 value of 0.9973. As illustrated in Fig. 6, the predicted reservoir water levels closely followed the observed values during the testing period. Furthermore, the scatter plot in Fig. 7 demonstrates a strong correlation between the observed and predicted reservoir water levels.

4.3 Quadratic Regression (QR) Results

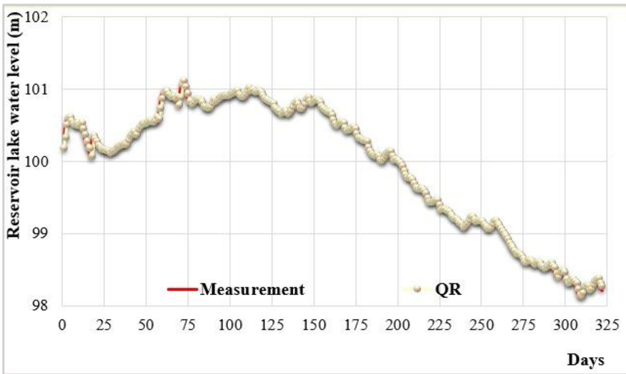

Fig. 8: Comparison of observed and predicted daily reservoir water levels from the QR model.

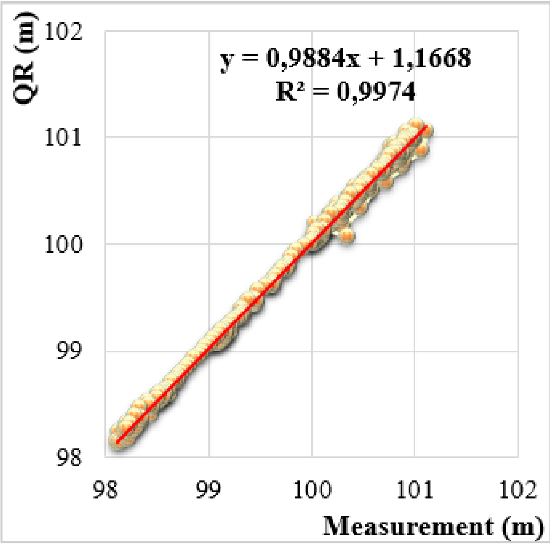

Fig. 9: Scatter plot of the observed versus predicted daily reservoir water levels for the Quadratic Regression (QR) model.

According to Table 1, the QR model achieved an MAE of 0.0313 m, an RMSE of 0.0450 m, and an R^2 value of 0.9974. As shown in Fig. 8, the predicted reservoir water levels closely followed the observed values throughout the testing period. Furthermore, the scatter plot in Fig. 9 shows strong agreement between the observed and predicted reservoir water levels. Compared with the MLR and IR models, the QR model exhibited slightly better prediction performance, as indicated by its lower error values and higher coefficient of determination.

4.4 Long Short-Term Memory (LSTM) Results

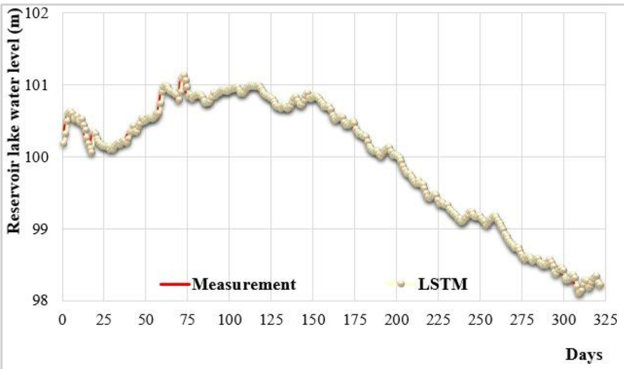

Fig. 10: Comparison between the observed and predicted daily reservoir water levels from the LSTM model.

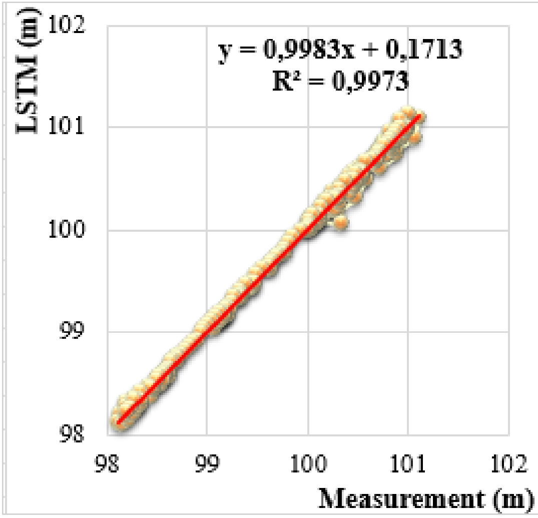

Fig. 11: Scatter plot of the observed versus predicted daily reservoir water levels for the LSTM model.

According to Table 1, the LSTM model achieved an MAE of 0.0326 m, an RMSE of 0.0460 m, and an R^2 value of 0.9973. As shown in Fig. 10, the predicted reservoir water levels closely followed the observed values throughout the testing period. Furthermore, the scatter plot in Fig. 11 shows strong agreement between the observed and predicted reservoir water levels. Overall, the LSTM model provided reliable and accurate predictions of daily reservoir water levels.

V. CONCLUSION

In this study, daily reservoir water levels of Thurmond Reservoir, located on the Savannah River between the states of South Carolina and Georgia, USA, were predicted using daily average relative humidity (RH), air temperature (T), precipitation (P), and one-day lagged reservoir water level [RWL(t−1)]. A total of 1608 daily observations collected between 2017 and 2023 were used, with 1286 (80%) allocated to training and 322 (20%) to testing. The predictive performance of the Multiple Linear Regression (MLR), Interaction Regression (IR), Quadratic Regression (QR), and Long Short-Term Memory (LSTM) models was evaluated using the Coefficient of Determination (R^2), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE).

The results showed that all developed models successfully predicted daily reservoir water levels with high accuracy. Among the evaluated models, the QR model achieved the best performance, with an MAE of 0.0313 m, an RMSE of 0.0450 m, and an R^2 value of 0.9974. The IR and LSTM models also produced highly accurate predictions with similar performance, whereas the MLR model yielded satisfactory results despite its relatively lower prediction accuracy. These findings demonstrate that both classical regression and deep learning approaches can be effectively applied to daily reservoir water level prediction. Furthermore, the proposed methodology may provide useful support for reservoir operation, water resources planning, and sustainable water resources management.

ACKNOWLEDGEMENTS

The data used in this study were obtained from the United States Geological Survey (USGS). The authors gratefully acknowledge the USGS [23] for providing open-access data.

REFERENCES

[1] UNESCO. (2024). United Nations World Water Development Report 2024: Water for Prosperity and Peace.

[2] Salas, J. D., Delleur, J. W., Yevjevich, V., & Lane, W. L. (1980). Applied Modeling of Hydrologic Time Series. Water Resources Publications.

[3] Bazartseren, B., Hildebrandt, G., & Holz, K. P. (2003). Short-term water level prediction using neural networks and neuro-fuzzy approach. Neurocomputing, 55(3-4), 439-450.

[4] Unes, F., Demirci, M., Tasar, B., Kaya, Y., & Varçin, H. (2019). Modeling of dam reservoir volume using generalized regression neural network, support vector machines and M5 decision tree models. Applied Ecology and Environmental Research, 17(3).

[5] Unes, F., Gumuscan, F.G., Demirci, M. (2017). Prediction of dam reservoir volume fluctuations using adaptive neuro fuzzy approach. European Journal of Engineering and Natural Sciences, 2(1), 144-148.

[6] Ünes, F. (2010). Dam reservoir level modeling by neural network approach: A case study. Neural Network World, 20(4), 461.

[7] Taşar, B., Unes, F., & Varcin, H. (2019). Prediction of the Rainfall–Runoff Relationship Using Neuro-Fuzzy and Support Vector Machines. Aerul si Apa. Componente ale Mediului, 237-246.

[8] Üneş, F., & Demirci, M. (2015). Generalized regression neural networks for reservoir level modeling. International Journal of Advanced Computational Engineering and Networking, 3(8), 81-84.

[9] Kaya, Y. Z., Mamak, M., & Unes, F. (2016). Evapotranspiration prediction using M5T data mining method. International Journal of Advanced Engineering Research and Science, 3(12), 236973.

[10] Barzegar, R., Raei, E., & Adamowski, J. (2026). Explaining Great Lakes water level variability through interpretable ensemble machine learning. Science of The Total Environment, 1013, 181302.

[11] Ojugo, A. A., Ejeh, P. O., Odiakaose, C. C., Eboka, A. O., & Emordi, F. U. (2026). Predicting rainfall runoff in Southern Nigeria using a fused hybrid deep learning ensemble. International Journal of Informatics and Communication Technology (IJ-ICT), 13(1), 108-115.

[12] Sharma, A., Shankar, V., & Thakur, P. K. (2026). Performance Evaluation of Metric and Parametric Models through Rainfall-Runoff Modelling of Sub-Himalayan Catchment. Water Resources, 53(1), 1-13.

[13] Özgür, A., & Yamaç, S. S. (2026). Modelling of daily reference evapotranspiration using deep neural network in different climates. Journal of Applied Water Engineering and Research, 14(2), 219–236. https://doi.org/10.1080/23249676.2025.2606417

[14] Raimondi, G., Bonato, J., & Maucieri, C. (2026). Machine learning models for reference evapotranspiration estimation in new locations under data-limited conditions in Northeastern Italy. European Journal of Agronomy, 178, 128132.

[15] Güzel, H., Üneş, F., Erginer, M., Kaya, Y. Z., Taşar, B., Erginer, İ., & Demirci, M. (2023). A comparative study on daily evapotranspiration estimation by using various artificial intelligence techniques and traditional regression calculations.

[16] Ozdemir, S., Yaqub, M., & Yildirim, S. O. (2023). A systematic literature review on lake water level prediction models. Environmental Modelling & Software163,105684. https://doi.org/10.1016/j.envsoft.2023.105684

[17] Zhu, S., Ji, Q., Ptak, M., Sojka, M., Keramatfar, A., Chau, K. W. ve Band, S. S. (2023). Daily water-level forecasting for multiple Polish lakes using multiple data-driven models. The Geographical Journal, 189(2), 357–369.

[18] USACE, U.S. Army Corps of Engineers. (2022). J. Strom Thurmond Dam and Lake.

[19] Google Earth. (2025). Google Earth. https://earth.google.com/

[20] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[21] Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. Neural Computation, 12(10), 2451–2471. https://doi.org/10.1162/089976600300015015

[22] Kılıç, D. K., & Uğur, Ö. (2023). Hybrid wavelet-neural network models for time series. Applied Soft Computing,144,110469. https://doi.org/10.1016/j.asoc.2023.110469

[23] U.S. Geological Survey. (2025). USGS Water Data for the Nation: U.S. Geological Survey National Water Information System (NWIS). https://doi.org/10.5066/F7P55KJN