

Article

Not peer-reviewed version

Hybrid Conv1D–LSTM Modelling of Short-Term Reservoir Water-Level Dynamics for Scenario-Based Operational Analysis

[Jelena Markovic Brankovic](#)*, [Milica Marković](#), [Bojan Branković](#)

Posted Date: 18 March 2026

doi: 10.20944/preprints202603.1350.v1

Keywords: Reservoir modelling;CNN–LSTM;Conv1D;water-level dynamics;scenario-based simulation;hydrological time series;operational water management.



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Hybrid Conv1D–LSTM Modelling of Short-Term Reservoir Water-Level Dynamics for Scenario-Based Operational Analysis

Jelena Marković Branković ^{1,*}, Milica Marković ¹ and Bojan Branković ²

¹ Faculty of Civil Engineering and Architecture, University of Nis, Serbia

² Adesso Austria GmbH, Wien, Austria

* Correspondence: jelena.markovic.brankovic@gaf.ni.ac.rs

What are the main findings?

- Conv1D–LSTM model reproduces short-term reservoir water levels.
- Optimal input window identified as 24 time steps.
- GELU activation improves stability over longer sequences.

What are the implications of the main findings?

- Validation stability is more important than minimal test error.
- Framework supports scenario-based operational analysis.
- Approach complements physically based reservoir models.

Abstract

Accurate representation of short-term reservoir water-level dynamics is essential for operational analysis and scenario-based assessment under prescribed inflow–outflow conditions. In many practical applications, physically based modelling is limited by incomplete process knowledge, unavailable boundary conditions, or insufficient temporal resolution of input data. This study presents a data-driven framework for hourly conditional simulation of reservoir water level based on a hybrid Conv1D–LSTM architecture. The model learns nonlinear relationships among hydraulic forcing, operational control, and system state from historical observations, and is evaluated in a recursive multi-step simulation (rollout) mode to reflect its intended use and capture error accumulation over time. A systematic analysis of input sequence length and activation function is performed to identify a robust model configuration. On the test set, the selected configuration ($L = 24$, GELU) achieved RMSE = 0.1057 m, MAE = 0.0881 m, and $R^2 = 0.972$ in rollout evaluation. The proposed framework is designed for scenario-based simulation rather than one-step deterministic forecasting, enabling rapid operational screening of alternative inflow–outflow regimes. Unlike many previous studies that emphasize one-step predictive accuracy, this work explicitly assesses model stability in recursive multi-step simulation, which is more relevant for reservoir scenario analysis.

Keywords: reservoir modelling; CNN–LSTM; Conv1D; water-level dynamics; scenario-based simulation; hydrological time series; operational water management

1. Introduction

Accurate representation of short-term reservoir water-level dynamics is fundamental for operational water management, risk assessment, and system understanding, particularly in the context of increasingly available high-resolution operational monitoring data [1]. In regulated reservoirs, water levels respond to a complex combination of hydrological forcing, operational control decisions, and seasonal variability [2,3]. These responses are inherently nonlinear and temporally dependent, reflecting cumulative storage effects, delayed system reactions, and regime

shifts induced by both natural and anthropogenic drivers [4]. Reliable modelling of these dynamics is therefore essential for short-term operational analysis and scenario-based evaluation [1].

Traditionally, reservoir water levels have been simulated using physically based hydrological or hydrodynamic models [5]. While these models provide physically interpretable representations of system behaviour, their application often requires detailed knowledge of boundary conditions, hydraulic parameters, and process formulations [5]. In practical operational settings, complete process information is frequently unavailable, and input data may be limited in temporal resolution [6]. These limitations motivate the development of data-driven modelling approaches that can complement physically based frameworks under operational constraints [7].

Recent advances in machine learning have enabled the application of deep neural networks to hydrological time-series modelling [8]. In particular, recurrent neural networks and Long Short-Term Memory (LSTM) variants have demonstrated significant capability in representing nonlinear system behaviour by effectively addressing the vanishing gradient problem [9,10]. This property is especially important for reservoir systems, where current water levels depend not only on instantaneous inflow but also on accumulated prior states [11]. Recent research further suggests that hybrid architectures, such as CNN-LSTM, can efficiently extract informative features while balancing representational capacity and computational efficiency [12].

Most existing data-driven applications focus on deterministic forecasting under unknown boundary conditions [13]. However, in many reservoir-management contexts, future inflow and operational strategies are prescribed as alternative scenarios, which shifts the modelling objective from forecasting toward conditional simulation. In such cases, the practical value of a model depends not only on one-step predictive accuracy but also on its stability during recursive multi-step use. Nevertheless, most prior studies primarily optimize one-step forecasting accuracy, whereas the present study evaluates and selects models for stable multi-step conditional (scenario-based) simulation in recursive rollout.

Existing studies have explored a wide range of enhancements in data-driven reservoir and water-level modelling, including hybrid workflows aimed at improving data integrity [14], physics-ML integration for large-scale daily simulations [15], and automated or optimization-based model configuration in smaller reservoirs [16,17]. Additional efforts have combined learning models with filtering or assimilation techniques to refine real-time estimates [18], applied signal-decomposition and metaheuristic tuning to reduce prediction errors [19], or used optimization schemes to improve scenario selection and stability in nonlinear settings [20]. Deep learning has also been evaluated in critical flood-season regimes, where peak tracking is particularly important [21]. Across these research directions, a recurring operational issue is multi-step stability, because recursive application can amplify one-step errors over time [16,22].

This study addresses these challenges by developing a hybrid Conv1D-LSTM framework designed explicitly for scenario-based simulation of short-term reservoir water-level dynamics in the Bovan Reservoir. The proposed architecture combines a one-dimensional convolutional layer for the automatic extraction of local temporal features with an LSTM component that captures longer-term dependencies and delayed system responses. Rather than selecting models solely based on a single best test score, the study systematically investigates the influence of input sequence length (look-back period) and activation function on model stability, robustness, and generalization within a recursive rollout process and across validation windows [12]. In this way, the study emphasizes performance characteristics that are directly relevant to operational use rather than relying exclusively on conventional one-step evaluation results.

Objectives and Contributions

The *objectives* of this study are to:

- develop a data-driven model for hourly simulation of reservoir water level under prescribed inputs (inflow and outflow), intended for operational scenario analyses;

- establish a model verification procedure based on multi-step simulation to assess whether the model is usable as a simulator, particularly with respect to stability and error accumulation over time;
- analyse the effects of input window length and activation function on accuracy and stability, and select a configuration suitable for operational use;
- provide recommendations for applying the model to typical reservoir-management tasks (e.g., checking compliance with operational water-level limits under different inflow/outflow regimes), including clearly stated limitations.
- The *main contributions* of this study are:
 - a clear distinction between forecasting and scenario-based simulation in the context of reservoir operations, clarifying how metrics and results should be interpreted [4,7];
 - the introduction of a multi-step simulation-based evaluation procedure that directly tests model stability and error accumulation in the intended mode of use;
 - a model selection criterion that, in addition to accuracy, incorporates performance stability across multiple validation periods;
 - a systematic analysis of key hyperparameters (input window length and activation function) and the identification of a robust configuration for hourly water-level simulation;
 - a demonstration based on operational SCADA measurements, explicitly considering data-quality issues (interruptions and missing values) and their implications for the reliability of scenario-based analyses in support of operational decision-making.

2. Materials and Methods

2.1. Study Area and Data

2.1.1. Study Reservoir: Bovan (Aleksinac, Serbia)

The study was conducted for the Bovan Reservoir, formed by the construction of the Bovan dam in the vicinity of Aleksinac (southeastern Serbia). The reservoir is multipurpose and is used for municipal water supply for Aleksinac and surrounding settlements, flood protection, sediment retention, low-flow augmentation, hydropower generation, and irrigation. Due to these functions, the reservoir represents an important regional infrastructure asset for water supply and water-regime management.

The catchment is characterized by a mixed rainfall-snow hydrological regime, with peak inflows typically occurring during spring and early summer. Reservoir operation is primarily governed by water-supply requirements and flood-protection constraints, which influence the seasonal storage strategy

The operational regulation type corresponds to annual flow equalization, implying pronounced seasonal components in the water-level regime as well as a dependence on antecedent storage conditions (system “memory” effects). This characteristic is particularly relevant for sequential modelling at an hourly time step.

The main reservoir characteristics (based on the available project/operational information) are: gross storage $60 \times 10^6 \text{ m}^3$, active storage $41 \times 10^6 \text{ m}^3$, dead storage $3 \times 10^6 \text{ m}^3$, minimum operating level 243.00 m a.s.l., normal water level 252.50 m a.s.l., and maximum water level 261.50 m a.s.l. In practice, reservoir operation is conducted within operational zones defined by the dead storage, minimum operating level, and normal/maximum water levels. Inflow–outflow scenarios are evaluated in terms of maintaining the water level within the prescribed limits. Water-level dynamics follow the storage balance $\Delta S = \int (Q_{in} - Q_{out}) dt$, while changes in storage are translated into water-level variations through the elevation–storage relationship (H–V), which motivates the need for stable multi-step simulation.

Climatically, the area belongs to the temperate zone with pronounced temperate-continental characteristics, with seasonal variability reflected in inflow conditions and water-level dynamics.

2.1.2. Data Sources and Measured Variables

The data were obtained from the operational SCADA system, which collects and archives real-time measurements from the monitoring and control infrastructure of the Bovan dam. The SCADA system at Bovan has been in operational use only in the last few years and, in practice, exhibits frequent data gaps and availability issues (missing values and discontinuities in the time series). Therefore, the available dataset represents a typical real-world operational setting in which measurements exist but are not ideal. This circumstance is both a limitation and a key motivation of the study: the objective is to develop and evaluate a model that remains usable for scenario-based analyses even when the dataset is relatively short and affected by interruptions.

Although the observation period is relatively short for data-driven modelling, the hourly resolution provides a sufficiently large number of samples for training deep learning models. Nevertheless, the limited temporal coverage may restrict representation of rare extreme events.

SCADA systems are widely used for monitoring and control of industrial and municipal infrastructure, enabling continuous acquisition of sensor measurements and operational records. In this study, the analysis is based on hourly (1 h) data covering the period from 18 May 2021 to 20 October 2022. The following variables were used:

- Reservoir water level, 1 h;
- Inflow proxy, represented by discharge at the Žužkovac hydrological station, 1 h. The station is part of the RHMZ reporting network (South Morava basin). It was established in 1967, with a long-term discharge record (since 1967) and digital water-level recording since 2007, and the catchment area is 394 km². These characteristics make Žužkovac a representative indicator of inflow variability in the study catchment [23];
- Outflow (release), defined as the aggregated discharge across all measurement points where releases are recorded, 1 h;
- Water temperature, 1 h.

Discharge at the Žužkovac station is used as a representative indicator of inflow variability in the study basin. Any mismatch between the station cross-section and the actual inflow entering the reservoir is treated as part of real-world measurement and data uncertainty.

2.1.3. Data Preprocessing and Missing-Value Handling

All signals were aligned to an hourly time step (1 h). Missing values were handled using a two-level strategy: short gaps (up to several consecutive hours) were filled by linear interpolation to preserve continuity for sequence construction, whereas longer gaps were not interpolated. Instead, any training or evaluation windows containing such discontinuities were excluded. This approach reduces the risk of introducing artificial trends and supports a more realistic assessment of model performance under operational data conditions.

2.1.4. Train/Validation/Test Split (Time-Ordered)

The dataset was split chronologically (time-ordered) without any random shuffling to prevent information leakage from future observations into the past. The first 80% of the time series was used for model development (training), with the last 10% of this training period reserved as a validation subset. The final 20% of the full record was retained as an independent test set.

Hyperparameter selection and early stopping were performed exclusively on the validation subset derived from the training period.

The test set remained untouched during model development and was used only for the final performance assessment in the recursive multi-step (rollout) simulation mode.

Table 1. Variables used in the analysis.

Variable	Symbol	Unit	Temporal resolution
Reservoir water level	H	m a.s.l	1 h

Inflow proxy	Q_{in}	m^3/s	1 h
Aggregated outflow	Q_{out}	m^3/s	1 h
Water temperature	T_w	$^{\circ}\text{C}$	1 h

All variables were aligned to the hourly (1 h) time step. Missing values were handled using a two-level strategy: short gaps were filled by linear interpolation, while longer gaps were not interpolated and windows containing discontinuities were excluded from training and evaluation.

Table 2. Time coverage and dataset split.

Dataset	Period	Role in modelling
Full dataset	18 May 2021–20 October 2022	Hourly inputs
Model training (with an internal validation subset)	18 May 2021–16 July 2022	Model training and validation
Test period	16 July 2022–20 October 2022	Final performance evaluation

Early-stopping note. Early stopping was applied only on the validation subset extracted from the training period, while the test set was kept independent and used exclusively for final evaluation in rollout mode. This protocol enables configuration selection based on stability within the validation part of the training period, while reporting multi-step simulation performance on an unseen test window.

2.1.5. Data Preprocessing and Feature Engineering

All signals (water level, inflow, outflow, and water temperature) were aligned to an hourly time step (1 h). Missing values were handled using a two-level strategy: short gaps were filled by linear interpolation, while longer gaps were not interpolated and any windows containing discontinuities were excluded from training and evaluation. To enable the model to learn periodic patterns typical of reservoir operation (daily operational cycles and seasonal variability), additional time-related covariates were derived from the timestamp. Specifically, the features Hour, DayOfYear, and Weekday were constructed to encode intra-day, annual, and weekly periodicity.

This approach allows part of the unstructured seasonal and operational effects to be explicitly represented in the input, without introducing additional physical process variables.

Table 3. Derived temporal features used in the model.

Derived feature	Symbol	Range/type	Resolution	Purpose in the model
Hour of day	Hour	0–23	1 h	Intra-day periodicity and operational patterns
Day of year	DayOfYear	1–365 (366)	1 h	Seasonal variability (hydrological and thermal regime)
Day of week	Weekday	0–6	1 h	Weekday/weekend differences; demand and water-supply operation

The Weekday feature was introduced as a proxy for changes in demand and operational management between working days and weekends, which can influence abstraction patterns and, indirectly, releases.

3. Methodology and Experimental Design

3.1. Problem Formulation and Input Windowing

The modelling objective is to reproduce the hourly dynamics of reservoir water level in a simulation (scenario-based) setting using multivariate hourly inputs [10,11]. Let the input vector at time t be defined as:

$$\mathbf{x}_t = [H_t, Q_{in,t}, Q_{out,t}, T_{w,t}, Hour_t, DayOfYear_t, Weekday_t] \quad (1)$$

where H_t is the reservoir water level, $Q_{in,t}$ is the inflow proxy (Žužkovac station), $Q_{out,t}$ is the aggregated outflow (release), $T_{w,t}$ is water temperature, and the remaining components are time covariates derived from the timestamp. The target variable is the reservoir water level at the next time step, H_{t+1} .

Scaling and normalization. To improve numerical conditioning during training, the target water level was manually normalized by a constant reference level, $LEVEL_DIV = 252.5$, such that

$$H_t^* = \frac{H_t}{LEVEL_DIV} \quad (2)$$

Model outputs were rescaled to physical units (m a.s.l.) for reporting by $H_t = H_t^* \cdot LEVEL_DIV$. All remaining continuous hydraulic inputs (Q_{in}, Q_{out}, T_w) were standardized using z-score normalization (StandardScaler). The scaler parameters (mean and standard deviation) were fitted on the training period only and then applied unchanged to the validation and test periods to avoid information leakage. Derived temporal covariates (Hour, DayOfYear, Weekday) were used as numerical inputs.

For an input window of length L , the model input sequence at time t is formed as:

$$\mathbf{X}_t = [x_{t-L+1}, x_{t-L+2}, \dots, x_t] \quad (3)$$

and the model produces a one-step-ahead estimate:

$$\hat{H}_{t+1} = f(\mathbf{X}_t) \quad (4)$$

where $f(\mathbf{X}_t)$ is realized by a hybrid Conv1D–LSTM neural network. Unlike classical forecasting under unknown future inputs, this study treats Q_{in} , Q_{out} , and T_w as prescribed sequences (observed or scenario-defined) and uses the model as a conditional simulator of water-level response. Therefore, in addition to one-step performance, particular emphasis is placed on recursive behavior and error accumulation in multi-step simulation (described below).

3.2. Recursive Multi-Step Rollout

In scenario-based use, simulation is performed recursively: the predicted water level $\hat{H}_t$ is used as part of the system state for the next step, together with prescribed inputs $\mathbf{x}_t$ [12]. The procedure can be summarized as follows:

1. Initialization at the start of the simulation window using the observed H_t in the first step;
2. One-step prediction of $\hat{H}_{t+1}$ based on the input sequence of length L ;
3. Window update by shifting the input window forward by one-time step;
4. State replacement in rollout: after initialization, the observed H within the input window is replaced by the predicted value $\hat{H}$ for all subsequent steps, while Q_{in} , Q_{out} , and T_w remain prescribed (observed or scenario-defined) inputs;
5. Iteration of this procedure over the entire simulation period.

This evaluation reflects the intended use of the model as a simulator and enables assessment of stability and error accumulation over time, especially during periods of rapid inflow and/or release changes.

3.3. Model Architecture (Hybrid Conv1D–LSTM)

The proposed model is based on a hybrid Conv1D–LSTM architecture that combines local temporal feature extraction with modelling of longer-term dependencies within a single framework. The model input is a multivariate sequence of length L , where the water level is used both as an input state variable and as the target variable. This allows the network to learn the dependence of future water level on antecedent storage state, which is physically meaningful for reservoirs.

At time t , the input vector is defined in Equation 1., and the input window is $X_t = [x_{t-L+1}, \dots, x_t]$. The model output is a one-step estimate $\hat{H}_{t+1}$. The processing pipeline is:

$$X_t \rightarrow \text{Conv1D} \rightarrow \text{LSTM} \rightarrow \text{Dense} \rightarrow \hat{H}_{t+1}.$$

The final architecture therefore consists of the following sequence: Conv1D $\rightarrow$ activation $\rightarrow$ dropout $\rightarrow$ LSTM $\rightarrow$ dense output layer.

3.3.1. Convolutional Component (Conv1D)

The Conv1D layer performs convolution along the temporal dimension of the input sequence and is used to automatically extract local patterns, such as:

- rapid water-level changes associated with sudden inflow variations or release regime shifts,
- short-term oscillations and daily operational patterns,
- local trend structures across multiple input signals.

In this sense, the convolutional layer acts as a transformation (and partial filtering) of raw inputs into representations that are more suitable for learning sequential dependencies in the LSTM component, potentially reducing sensitivity to noise.

3.3.2. Recurrent Component (LSTM)

The LSTM layer models longer-term temporal dependencies through its gated memory mechanism, allowing information from past states to be retained over longer horizons [11]. This is particularly important for reservoirs due to system inertia and the fact that water level reflects cumulative effects of inflow, outflow, and antecedent storage. The LSTM component therefore supports representation of delayed response and slower dynamics that may not be captured by local patterns alone.

3.3.3. Output Layer (Dense)

The LSTM output is passed to a fully connected (Dense) layer that produces a scalar estimate $\hat{H}_{t+1}$. The model is trained to minimize the difference between $\hat{H}_{t+1}$ and the observed H_{t+1} . During evaluation, the recursive rollout regime is additionally used to assess stability and error accumulation over multi-step simulation.

The model was implemented as a Conv1D–LSTM hybrid with 64 convolutional filters (kernel size = 3) followed by an LSTM layer with 64 units. To reduce overfitting, dropout of 0.2 was applied. The model was trained using mean squared error (MSE) loss and the Adam optimizer (learning rate = 0.001) with a batch size of 32 for up to 200 epochs. Early stopping monitored the validation loss (val_loss) with a patience of 10 epochs, and the model corresponding to the lowest validation loss was retained for final evaluation. The main hyperparameters analysed were the input window length L and the activation function (ReLU, ELU, GELU, Swish, tanh), while other parameters were kept fixed to enable fair comparison across configurations.

3.4. Performance Metrics

Model performance was evaluated using standard error and agreement metrics for time series:

1. Mean Absolute Error (MAE) represents the average absolute deviation of simulated from observed water level and is expressed in meters; it is relatively robust to occasional larger errors:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |H_i - \hat{H}_i| \tag{5}$$

2. Root Mean Squared Error (RMSE) places higher weight on larger deviations and is informative during periods of rapid water-level changes [13]:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (H_i - \hat{H}_i)^2} \tag{6}$$

3. Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^N (H_i - \hat{H}_i)^2}{\sum_{i=1}^N (H_i - \bar{H})^2} \tag{7}$$

where $\bar{H}$ is the mean observed water level over the evaluation window. R^2 quantifies agreement in explained variance and is reported as a complementary indicator of reproduction quality. All metrics were computed over the test window in the scenario-based simulation setting, where $\hat{H}$ is obtained via recursive multi-step rollout.

3.5. Model Selection Criteria

Although the minimum test RMSE is a common selection criterion in forecasting tasks, in scenario-based reservoir simulation an isolated optimum on a single test segment may reflect sensitivity to a particular inflow/outflow regime. Therefore, model selection in this study was not based solely on minimal test error, but on a combination of:

- accuracy (RMSE, MAE, R^2);
- training/validation stability (consistent improvement without strong oscillations);
- robustness with respect to hyperparameter choices (reduced sensitivity to L and activation function);
- rollout behaviour, i.e., limited error accumulation during recursive simulation.

In other words, configurations showing consistent and stable performance were preferred, even if they were not the absolute best on a single test segment, because stability is critical for operational scenario testing.

3.6. Summary of Hyperparameter Experimental Design

A systematic grid evaluation was conducted over combinations of input sequence length $L \in \{12, 24, 30, 48, 72\}h$ and activation functions (ReLU, ELU, GELU, Swish, tanh). Other parameters (Conv1D filters and kernel size, LSTM units, dropout, and optimizer settings) were fixed as in Table 4 to ensure consistent comparison. The final configuration was selected as a compromise between accuracy and robustness in recursive multi-step simulation.

Table 4. Model configuration and parameters.

Component	Setting	Value
Input	Sequence length	$(L \in 12, 24, 30, 48, 72)h$
Input	Number of features per time step	$7((H, Q_{in}, Q_{out}, T_w, \text{Hour}, \text{DayOfYear}, \text{Wee})$
Conv1D	Number of filters	64
Conv1D	Kernel size	3
Activation function	Candidates	ReLU, ELU, GELU, Swish, tanh
LSTM	Number of units	64
Dropout	Dropout rate	0.2

Output	Dense layer	1 (linear output)
Optimizer	Type	Adam
Optimizer	Learning rate	0.001

4. Results

4.1. Overall Performance of the Selected Model

Within the test window (16 July 2022–20 October 2022), the selected configuration (L=24 with GELU activation) shows strong agreement with the observed water levels in the scenario-based simulation setting. As reported in Table 5, the model achieves RMSE = 0.1057 m and MAE = 0.0881 m, corresponding to typical deviations of approximately 10.6 cm (RMSE) and 8.8 cm (MAE). The high coefficient of determination ($R^2 = 0.972$) indicates that the model reproduces the dominant variability of the observed water-level time series. It should be emphasized that these metrics were obtained in a recursive rollout regime, where the simulated water level at time t is used as an input for predicting the next step ($t + 1$). Therefore, the reported performance reflects the model’s reproduction (simulation) capability under prescribed inputs (inflow, outflow, and water temperature), rather than classical one-step forecasting accuracy under uncertain future boundary conditions.

The persistence baseline yields substantially larger errors than the proposed model, confirming that the network learns meaningful system dynamics rather than merely exploiting water-level inertia.

Table 5. Overall performance of the selected configuration (rollout, 16 July 2022–20 October 2022).

Configuration	RMSE (m)	MAE (m)	R ²	Note
L = 24 + GELU	0.105726	0.088099	0.972138	Rollout (recursive multi-step simulation)

Table 6. Comparison with a persistence baseline (rollout, 16 July 2022–20 October 2022).

Model	MAE (m)	RMSE (m)	R ²
Conv1D–LSTM	0.0880	0.1057	0.9721
Persistence baseline (rollout, constant (H_{t0}))	1.0116	1.1575	–3.2322

LSTM-Only Baseline

To quantify the benefit of the convolutional feature-extraction stage, an LSTM-only baseline was trained by removing the Conv1D layer while keeping the remaining settings comparable [14]. In rollout evaluation over the test window, the LSTM-only model achieved RMSE = 0.217 m, MAE = 0.190 m, and $R^2 = 0.882$ (Table 7). The proposed Conv1D–LSTM model substantially improved performance (RMSE = 0.106 m, MAE = 0.088 m, $R^2 = 0.972$), corresponding to approximately a 51% reduction in RMSE and a 54% reduction in MAE. These results indicate that the convolutional component contributes to more effective extraction of local temporal patterns and improves stability in recursive multi-step simulation.

Table 7. Comparison between the proposed Conv1D–LSTM and an LSTM-only baseline (rollout evaluation, test window).

Model	RMSE (m)	MAE (m)	R ²
LSTM-only (without Conv1D)	0.2173	0.1898	0.8819
Proposed Conv1D–LSTM (L=24, GELU)	0.1057	0.0881	0.9721

4.2. Effect of Input Sequence Length L

Model performance shows a non-monotonic dependence on the input sequence length L , as reflected both in test metrics (Figure 1) and in validation stability indicators (Figure 2).

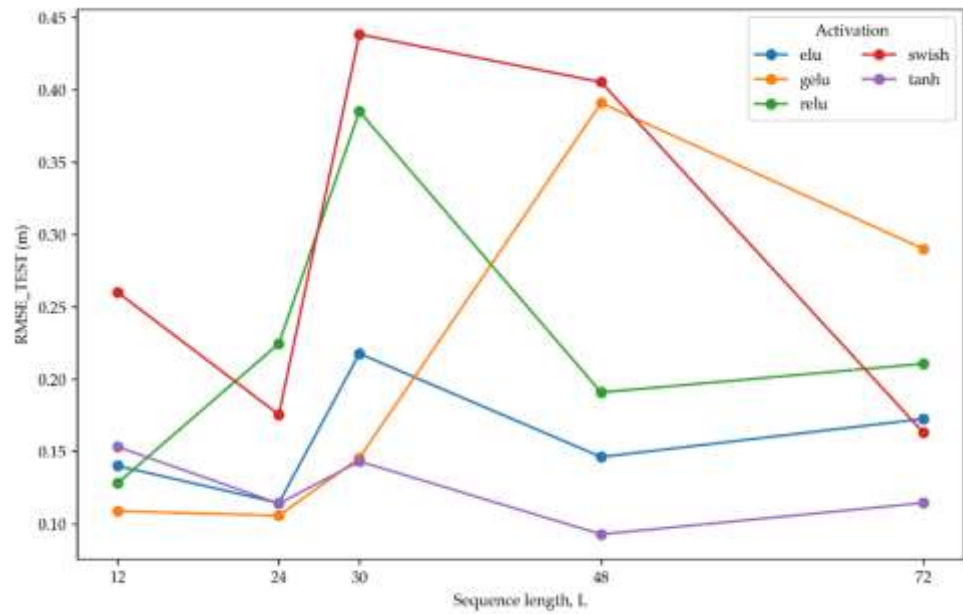


Figure 1. RMSE_test vs L .

For shorter sequences (e.g., $L = 12$ and $L = 24$), competitive values of $\text{RMSE}_{\text{test}}$ are obtained, indicating that the model can learn short-term dynamics effectively when the temporal context is limited. However, if L is too short, delayed response and storage (“memory”) effects may be insufficiently represented, which can become apparent in recursive rollout through increased sensitivity to rapid input changes and error propagation. As L increases to moderate values, the model gains sufficient context to represent both fast and slower components of water-level dynamics, often yielding a favourable compromise between accuracy and stability.

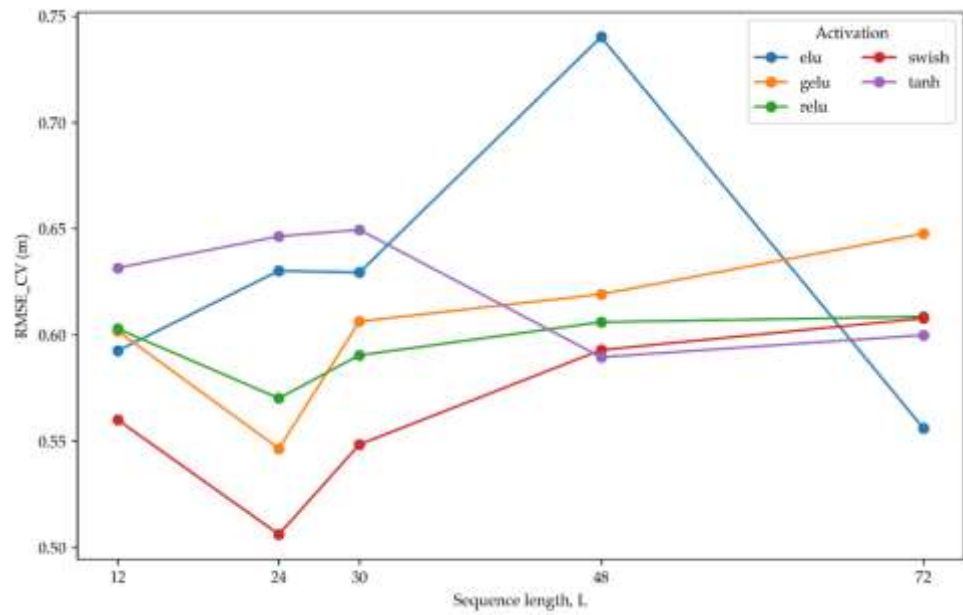


Figure 2. RMSE_cv vs L .

These results indicate that increasing the historical context beyond approximately one day does not provide additional predictive information but instead introduces redundancy that reduces model stability.

In contrast, extending the input window to larger values (e.g., $L = 48$ and especially $L = 72$) does not lead to systematic improvement: although some combinations may produce very low $\text{RMSE}_{\text{test}}$ (Figure 1), validation indicators (Figure 2) tend to deteriorate, suggesting higher variance, redundancy in historical information, and increased sensitivity to the specific regime of the evaluated period.

These results support the existence of an “optimal” context range in which accuracy and robustness are balanced; therefore, the choice of L in this study is based not only on minimal test error but also on validation stability, which is critical for scenario-based multi-step simulation.

4.3. Effect of Activation Function

The choice of activation function significantly affects accuracy and, more importantly, stability and robustness across combinations with the input sequence length L (Table 8).

The lowest isolated test RMSE in the grid evaluation was achieved by the combination $L = 48$ with tanh activation ($\text{RMSE}_{\text{test}} = 0.0928$ m; $R^2_{\text{test}} = 0.979$), but this configuration shows poor validation stability ($\text{RMSE}_{\text{CV}} = 0.694$ m), indicating sensitivity to the specific test-period regime and limited generalization. In contrast, GELU activation provides a more favourable compromise between accuracy and stability [15].

The configuration $L = 24$ with GELU achieves strong test performance ($\text{RMSE}_{\text{test}} = 0.1057$ m; $R^2_{\text{test}} = 0.972$) with comparatively better stability among the top candidates ($\text{RMSE}_{\text{CV}} = 0.656$ m).

This pattern suggests that a smoother nonlinearity (GELU) contributes to more stable learning and reduced sensitivity to hyperparameter choices compared to configurations that minimize test error but exhibit unstable validation behaviour.

Given the operational objective of using the model as a simulator, preference is therefore given to configurations with a better accuracy–stability balance.

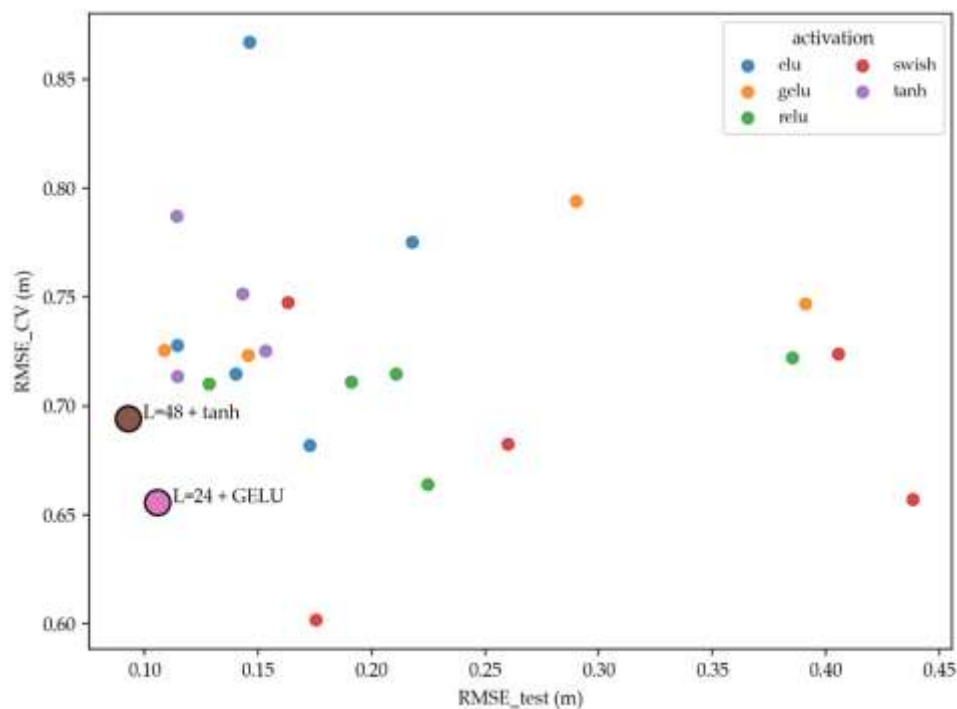


Figure 3. Accuracy vs Stability: $\text{RMSE}_{\text{test}}$ vs RMSE_{CV} .

The large difference between $RMSE_{test}$ and $RMSE_{CV}$ suggests that validation windows include different operational regimes and/or more dynamic episodes than the test window, increasing performance variability. Therefore, $RMSE_{CV}$ is used as an indicator of robustness rather than as a directly comparable metric to $RMSE_{test}$ from a single test segment.

Table 8. Top configurations by $RMSE_{test}$ with stability indicators.

Rank	L	Activation	RMSE_test (m)	R ² _test	RMSE_CV (m)
1	48	tanh	0.092834	0.978518	0.694146
2	24	GELU	0.105726	0.972138	0.655661
3	12	GELU	0.108810	0.970488	0.725610

4.4. Qualitative Comparison of Time Series (Observed vs. Simulated)

A qualitative comparison of observed and simulated water levels (Figure 4) shows that the model reproduces overall trends, the timing of maxima and minima, and the rates of increase and decrease.

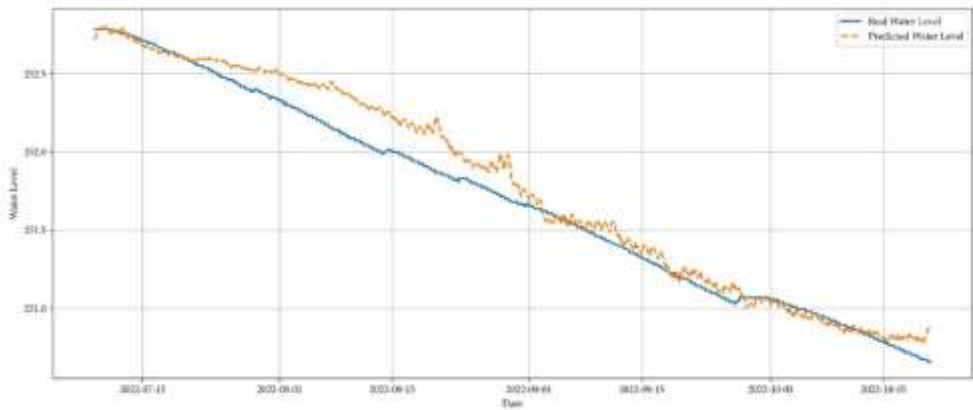


Figure 4. Observed vs Simulated water level.

The largest deviations occur during periods of rapid changes in inflow and/or outflow, where recursive rollout can lead to error accumulation. Nevertheless, the overall structure of the dynamics is preserved. Additionally, the residual time series (Figure 4) helps identify episodes with systematic deviations and assess whether errors exhibit trends or are predominantly episodic.

4.5. Error Analysis during Dynamic Periods (Event-Based)

A closer inspection of selected dynamic episodes—illustrated by the zoomed interval from 20 August 2022 to 30 August 2022 (Figure 5)—indicates that errors are mainly associated with phase shifts during rapid changes in the forcing signals, as well as occasional under- or overestimation of peaks.

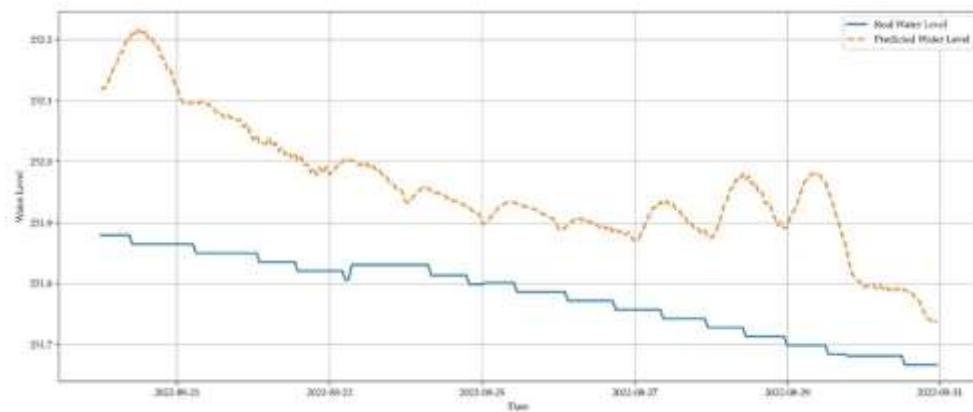


Figure 5. Zoomed view of observed and simulated water level (20 August 2022–30 August 2022).

This behaviour is expected in a recursive simulation setting, where small one-step errors can propagate over multiple steps, especially during fast transients. From an operational perspective, this analysis is important because it identifies conditions under which scenario simulations should be interpreted with additional caution (e.g., extreme inflow episodes or abrupt release regime changes). In such situations, it may be beneficial to consider a shorter recursive horizon or to perform additional checks using alternative scenarios.

5. Discussion

5.1. Interpretation of Results in the Context of Scenario-Based Simulation

The results show that the hybrid Conv1D–LSTM framework can reproduce the hourly water-level dynamics of the Bovan reservoir with a high degree of agreement in a scenario-based setting, i.e., under prescribed inputs (inflow, outflow, and water temperature).

Importantly, performance was not evaluated as a one-step forecast, but through recursive rollout, which better reflects the intended practical use of the model as a simulator for operational “what-if” analyses.

In this context, the selected configuration ($L=24 + \text{GELU}$) yields typical deviations on the order of centimetres while explaining a large fraction of the observed variance, indicating that the model captures dominant operational patterns and hydrological forcing. Relative to the operational water-level range and typical daily water-level variations, the achieved error is sufficiently small to support scenario-based comparison of operating regimes.

5.2. Effect of Sequence Length and an “Optimal” Temporal Context

The non-monotonic dependence of performance on the input window length L suggests the existence of a characteristic temporal horizon that provides sufficient information about antecedent conditions and forcing without introducing redundancy. Shorter windows may be adequate for capturing short-term oscillations but can miss part of the delayed response and storage (“memory”) effects of the reservoir. Conversely, overly long sequences can introduce redundant information and increase model variance, which in rollout mode manifests as greater sensitivity and reduced stability. The selected window $L = 24$ represents a practical compromise between capturing relevant context and maintaining stable generalization.

5.3. Activation Function and Validation Stability

The activation-function analysis demonstrates that the minimum test error alone is not a sufficient criterion for model selection. For example, the tanh configuration achieves the lowest RMSE in the test window, but shows markedly poorer validation stability, indicating sensitivity to the specific regime of the evaluated segment and limited robustness. In contrast, GELU provides a more

stable compromise between accuracy and robustness. In scenario-based simulation, such a compromise is more important than an isolated optimum, because operational use requires reliable behaviour across different inflow and operating regimes.

5.4. Errors during Dynamic Episodes and Error Accumulation in Rollout

Residual analysis and the event-based inspection indicate that the largest deviations are associated with periods of rapid changes in the forcing signals (inflow/outflow). In recursive simulation, small one-step errors can propagate and accumulate over multiple steps, producing phase shifts and occasional under- or overestimation of peaks. This behaviour is expected and is an important practical finding: the model is reliable for overall dynamics and trends, whereas fast transients warrant additional attention (e.g., targeted evaluation of critical episodes, additional input variables, or complementary checks).

5.5. Study Limitations

Although the results confirm good reproduction of hourly dynamics, several limitations should be noted. First, the model does not explicitly represent physical processes such as evaporation, infiltration, or detailed catchment-state dynamics; their influence is only indirectly reflected through inflow and the time covariates. Second, uncertainty is not quantified, as the framework is formulated deterministically [16,17]. Third, the observation period is relatively short and based on operational SCADA data affected by interruptions, which may limit generalization to rare and extreme regimes and highlights the need for additional verification on longer records when they become available.

5.6. Mass-Balance Consistency

To evaluate the physical consistency of the recursive simulation, a volumetric mass-balance check was performed over the test window [18]. The residual storage imbalance was computed as the difference between the simulated storage change and the integrated inflow–outflow difference. The mean residual was 13,160 m³, corresponding to approximately 0.03% of the useful storage volume (41×10^6 m³), indicating no systematic volumetric drift. The standard deviation of the residual was 13,328 m³. Maximum and minimum residuals ($\approx 4.89 \times 10^5$ m³ and -2.58×10^5 m³) were observed during rapid dynamic transitions and correspond to about 1.2% of useful storage. Overall, the residual distribution remains centred around small values relative to reservoir capacity, confirming that the recursive rollout does not introduce significant cumulative mass inconsistency at the operational scale.

5.7. Practical Implications and Future Work

From a practical perspective, the proposed model can serve as an efficient tool for rapid testing of operational scenarios (alternative release policies and prescribed inflow trajectories), particularly when physically based models are not available or are too slow for iterative analyses. Typical applications include rapid screening of how alternative release regimes affect compliance with operational level limits, as well as scenario-based response analysis under prescribed inflow waves (e.g., “what-if” episodes) for operational planning [19,20]. Future work may include extending the dataset as additional operational records become available, introducing uncertainty estimation (e.g., bootstrap or MC dropout), and incorporating additional inputs to better represent fast transients.

6. Conclusions

This study developed a data-driven framework based on a hybrid Conv1D–LSTM architecture for hourly scenario-based simulation of water level in the Bovan reservoir under prescribed inputs (inflow, outflow, and water temperature). The model was evaluated in a recursive multi-step (rollout) mode, explicitly accounting for error accumulation over time and aligning performance assessment with the intended operational use.

Results show that the selected configuration (L=24 + GELU) achieves high agreement with observed water levels over the test window, with RMSE = 0.1057 m ($\approx$ 10.6 cm), MAE = 0.0881 m ($\approx$ 8.8 cm), and $R^2 = 0.972$, where the metrics are obtained in recursive multi-step simulation. The systematic hyperparameter analysis confirmed that performance does not improve monotonically with increasing input length: very short windows may be insufficient to capture delayed response and storage effects, whereas overly long windows can introduce redundant information and increase variance, which is reflected in reduced stability of the recursive simulation.

The study further showed that minimal test error alone is not a sufficient selection criterion, because some configurations (e.g., with tanh activation) can yield low test RMSE while exhibiting poorer validation stability.

The main contributions of this work are:

- a clear distinction between scenario-based (conditional) simulation and classical forecasting in the context of reservoir operation;
- rollout-based evaluation as a criterion aligned with the real mode of model use in operational “what-if” analyses;
- a model selection approach that incorporates validation stability in addition to accuracy, supporting robust application across varying inflow and operating regimes;
- demonstration of applicability on operational hourly SCADA time series under realistic constraints of data availability and quality.

Practically, the proposed model can support rapid assessment of the effects of alternative release regimes and prescribed inflow scenarios on maintaining water levels within operational limits, particularly when detailed physically based models are unavailable or when many scenarios must be evaluated iteratively.

Key limitations include the relatively short observation period and the deterministic formulation without explicit uncertainty quantification.

Future work may therefore focus on expanding the dataset, incorporating uncertainty estimation (e.g., bootstrap/MC dropout), and adding inputs that better describe fast dynamic transitions.

Author Contributions: Conceptualization, J.M.B. and B.B.; methodology, J.M.B. and B.B.; software, B.B.; validation, J.M.B., B.B. and M.M.; formal analysis, J.M.B.; investigation, J.M.B.; resources, M.M.; data curation, B.B.; writing—original draft preparation, M.M.; writing—review and editing, M.M.; visualization, B.B.; supervision, J.M.B. and M.M.; project administration, J.M.B.; funding acquisition, M.M.

Funding: This research received no external funding.

Data Availability Statement: Operational SCADA data used in this study are not publicly available due to data ownership and operational/security restrictions.

Sustainable Development Goals: Sustainable Cities and Communities.

Acknowledgments: This research was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, under the Agreement on Financing the Scientific Research Work of Teaching Staff at the Faculty of Civil Engineering and Architecture, University of Niš - Registration number: 451-03-34/2026-03/200095 dated 05/02/2026.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Conv1D	One-dimensional convolution
LSTM	Long Short-Term Memory
CNN	Convolutional Neural Network

SCADA	Supervisory Control and Data Acquisition
ReLU	Rectified Linear Unit
ELU	Exponential Linear Unit
GELU	Gaussian Error Linear Unit
Swish	Swish activation function
tanh	Hyperbolic tangent
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
R^2	Coefficient of Determination
CV	Cross-validation / validation-window indicator
H–V	Elevation–storage relationship
Q_{in}	Inflow (proxy)
Q_{out}	Outflow / release
T_w	Water temperature

References

1. Lee, Y.K.; Hong, S.; Kim, S.W. Monitoring of Water Level Change in a Dam from High-Resolution SAR Data. *Remote Sensing* **2021**, *13*. <https://doi.org/10.3390/rs13183641>
2. Yang, G.; Block, P. Water Sharing Policies Conditioned on Hydrologic Variability to Inform Reservoir Operations. *Hydrology and Earth System Sciences* **2021**, *25*, 3617–3634. <https://doi.org/10.5194/hess-25-3617-2021>
3. Costa, A.; Anghileri, D.; Molnár, P. Hydroclimatic Control on Suspended Sediment Dynamics of a Regulated Alpine Catchment: A Conceptual Approach. *Hydrology and Earth System Sciences* **2018**, *22*, 3421–3434. <https://doi.org/10.5194/hess-22-3421-2018>
4. Coutinho, R.M.; Kraenkel, R.A.; Prado, P.I. Catastrophic Regime Shift in Water Reservoirs and São Paulo Water Supply Crisis. *PLoS ONE* **2015**, *10*(9). <https://doi.org/10.1371/journal.pone.0138278>
5. Jin, Y.; Liu, D.; Huang, J.J. Forecasting of Reservoir Water Level by Remote Sensing and Deep Learning. *Research Square* **2024**, *9*. <https://doi.org/10.21203/rs.3.rs-3984208/v1>
6. Sekulić, G.; Ivković, M.; Ćipranić, I. Modelling of Hydrological Processes in the Catchment Area of Lake Skadar. *Technical Gazette* **2017**, *24*, 427-434. <https://doi.org/10.17559/TV-20151120230958>
7. Mohammed, S.J.; Zubaidi, S.L.; Ortega-Martorell, S.; Al-Ansari, N.; Ethaib, S.; Hashim, K. Application of Hybrid Machine Learning Models and Data Pre-Processing to Predict Water Level of Watersheds: Recent Trends and Future Perspective. *Cogent Engineering* **2022**, *9* (1).<https://doi.org/10.1080/23311916.2022.2143051>
8. Zhao, X.; Wang, H.; Bai, M.; Xu, Y.; Dong, S.; Rao, H.; Ming, W. A Comprehensive Review of Methods for Hydrological Forecasting Based on Deep Learning. *Water* **2024**, *16*(10). <https://doi.org/10.3390/w16101407>
9. Roudbari, N.S.; Punekar, S.R.; Patterson, Z.; Eicker, U.; Poullis, C. From Data to Action: Flood Forecasting Leveraging Graph Neural Networks and Digital Twin Visualization. *Scientific Reports* **2024**, *14*. <https://doi.org/10.1038/s41598-024-68857-y>
10. Sahraei, A.; Houska, T.; Breuer, L. Deep Learning for Isotope Hydrology: The Application of Long Short-Term Memory to Estimate High Temporal Resolution of the Stable Isotope Concentrations in Stream and Groundwater. *Frontiers in Water* **2021**, *3*. <https://doi.org/10.3389/frwa.2021.740044>
11. Denaro, S.; Anghileri, D.; Giuliani, M.; Castelletti, A. Informing the Operations of Water Reservoirs over Multiple Temporal Scales by Direct Use of Hydro-Meteorological Data. *Advances in Water Resources* **2017**, *103*, 51-63. <https://doi.org/10.1016/j.advwatres.2017.02.012>
12. Li, H.; Zhang, L.; Yao, Y.; Zhang, Y. Prediction of Reservoir Water Levels via an Improved Attention Mechanism Based on CNN – LSTM. *Applied Intelligence* **2025**, *55*. <https://doi.org/10.1007/s10489-025-06393-6>
13. Özdoğan-Sarıkoç, G.; Dadaşer-Çelik, F. Physically Based vs. Data-Driven Models for Streamflow and Reservoir Volume Prediction at a Data-Scarce Semi-Arid Basin. *Environmental Science and Pollution Research* **2024**, *31*. <https://doi.org/10.1007/s11356-024-33732-w>

14. Liu, M.; Zhang, M.; Zhang, P.; Wang, G.; Chen, X.; Zhang, H. Water Level Prediction Model Based on Blockchain and LSTM. *Journal of Intelligent and Fuzzy Systems* **2023**, *46*, 2371-2380. <https://doi.org/10.3233/JIFS-231411>
15. Xie, M.; Shan, K.; Zeng, S.; Wang, L.; Gong, Z.; Wu, X.; Yang, B.; Shang, M. Combined Physical Process and Deep Learning for Daily Water Level Simulations across Multiple Sites in the Three Gorges Reservoir, China. *Water* **2023**, *15*. <https://doi.org/10.3390/w15183191>
16. Li, H.; Zhang, L.; Zhang, Y.; Yao, Y.; Wang, R.; Dai, Y. Water-Level Prediction Analysis for the Three Gorges Reservoir Area Based on a Hybrid Model of LSTM and Its Variants. *Water* **2024**, *16*. <https://doi.org/10.3390/w16091227>
17. He, R.; Jia, W.; Qian, Z. Deriving Implicit Optimal Operation Rules for Reservoirs Based on TgLSTM. *Water* **2025**, *17*. <https://doi.org/10.3390/w17142059>
18. Fu, J.-C.; Su, M.-P.; Liu, W.-C.; Huang, W.-C.; Liu, H.-M. Water Level Forecasting Combining Machine Learning and Ensemble Kalman Filtering in the Danshui River System, Taiwan. *Water* **2024**, *16*. <https://doi.org/10.3390/w16233530>
19. Malekpoor, M.M.; Malekpoor, H. Reservoir Water Level Forecasting Using Wavelet Support Vector Regression (WSVR) Based on Teaching Learning-Based Optimization Algorithm (TLBO). *Soft Computing* **2022**, *26*, 8897-8909. <https://doi.org/10.1007/s00500-022-07296-1>
20. Sammen, S.S.; Ehteram, M.; Sheikh Khozani, Z.; Sidek, L.M. Binary Coati Optimization Algorithm- Multi-Kernel Least Square Support Vector Machine-Extreme Learning Machine Model (BCOA-MKLSSVM-ELM): A New Hybrid Machine Learning Model for Predicting Reservoir Water Level. *Water* **2023**, *15*. <https://doi.org/10.3390/w15081593>
21. Mao, X.; Xiong, B.; Li, T.; Luo, X.; Yao, Z.; Li, J.; Huang, Y. Short Term Prediction of Water Level Based on Deep Learning during the Flood Season, in the Downstream Area of The Three Gorges Reservoir. *Natural Hazards* **2024**, *120*, 14259–14278. <https://doi.org/10.1007/s11069-024-06772-1>
22. Rohli, E.; Woolsey, N.; Sathiaraj, D. Near-Term Forecasting of Water Reservoir Storage Capacities Using Long Short-Term Memory. *Environmental Data Science* **2023**, *2*. <https://doi.org/10.1017/eds.2023.25>
23. Republic Hydrometeorological Service of Serbia (RHMZ). Surface Water Station: Žučkovac (hm_id=47580). Available online: https://hidmet.gov.rs/latin/hidrologija/povrsinske/pov_stanica.php?hm_id=47580 (accessed on 07 March 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.