Knowledge-based Engineering and Sciences (2023), **4**, (2), 73–88
doi:10.51526/kbes.2023.4.2.73-88

RESEARCH PAPER

A Temporal Fusion Transformer Deep Learning Model for Long-Term Streamflow Forecasting: A Case Study in the Funil Reservoir, Southeast Brazil

Geane Fayer,[1,2] Nelson Bolotari,[3,4] Felix Miranda,[5] Jussara S. Ferreira,[5] Renan R. C. da Silva,[5] Vinícius B. R. Cândido,[5] Marcel P. Andrade,[5] Marconi Morais,[5] Celso B. M. Ribeiro,[5] Priscila V. Z. Capriles,[5] and Leonardo Goliatt[*,5]

[1]ArcelorMittal, Rodovia BR 040, km 769, Barreira do Triunfo, Juiz de Fora, 36092-901, MG, Brazil
[2]Rio de Janeiro State University, Rua São Francisco Xavier, 524, Maracanã, Rio de Janeiro, 20550-900, RJ, Brazil
[3]ArcelorMittal, Avenida Getulio Vargas, 100, Centro Industrial, João Monlevade, 35930-900, MG, Brazil
[4]Federal University of Itajubá, Rua Irmã Ivone Drumond, 200, Distrito Industrial II, Itabira, 35903-087, MG, Brazil
[5]Federal University of Juiz de Fora, Rua José Lourenço Kelmer, S/N, Campus Universitario,Juiz de Fora,36036-900, MG, Brazil
[*]Corresponding author. Email: leonardo.goliatt@ufjf.br

(Received 13 July 2023; revised 23 August 2023; accepted 27 August 2023; first published online 31 August 2023)

Abstract

Water reservoirs play a critical role in water resource management systems, serving various purposes such as water supply, hydropower generation, and flood control. Accurate long-term streamflow predictions are essential for the efficient operation and planning of reservoirs, enabling water managers to anticipate changes in water availability, optimize reservoir storage, and make informed decisions about water allocation and infrastructure management. However, the increasing variability and uncertainty in hydrological processes due to climate change and anthropogenic activities necessitate the development of robust and precise prediction models. Temporal Fusion Transformer (TFT) models have emerged as a promising approach for hydrological forecasting, leveraging deep learning, time series analysis, and attention mechanisms to capture complex temporal dependencies and provide accurate predictions. This study employs TFT as a surrogate model to simulate the streamflow upstream of the Funil reservoir. A comparison was performed among the models Seasonal Naive, AutoARIMA, Theta method, and Deep ARIMA. The TFT model has the lowest MAE (70.88 m^3/s) and RMSE (121.66 m^3/s) of all models, which indicates that it is the most accurate one. The TFT model also has the highest NSE (0.43) and coefficient of determination (0.79), which indicates that it is the most promising model for capturing the actual streamflow patterns. The TFT model effectively captures the intricate spatiotemporal patterns and dependencies in the streamflow data and accurately predicts streamflow upstream of the Funil reservoir, capturing seasonal patterns and long-term trends. This can help water managers make informed decisions about reservoir operations and management.

Keywords: Streamflow simulation; Deep learning; Temporal fusion transformer; Funil reservoir.

1. Introduction

1.1 Research background

Assessing and managing water quantities in reservoirs are important for sustainable water resource planning and management. Water reservoirs are crucial storage systems that ensure reliable water supply for various sectors, including agriculture, industry, and domestic use [1, 2]. Accurately estimating and monitoring water quantities in reservoirs are essential for optimizing water allocation, reservoir operations, and drought management strategies. Understanding the volume and availability of water in reservoirs allows for efficient planning and utilization of water resources [3]. Additionally, water quantity information aids in flood control and mitigation, as reservoirs can store excess water during heavy rainfall, reducing the risk of downstream flooding [4]. Furthermore, assessing reservoir water quantities is vital for environmental management and maintaining ecological balance [5], enabling the preservation of adequate water levels to support aquatic ecosystems and safeguard biodiversity. Therefore, a comprehensive understanding of water quantities in reservoirs is critical for effective water resource planning, sustainable development, and the resilience of water systems in the face of increasing water demand and climate change [6, 7].

Physics-based hydrologic models (PBHM) [8, 9, 10, 11] provide a mechanistic understanding of river flow dynamics, enabling comprehensive analysis and prediction. However, their implementation requires careful data collection, parameter estimation, and consideration of model limitations to ensure accurate and reliable simulations [12]. PBHMs provide a mechanistic representation of the hydrologic processes, incorporating factors such as rainfall, evapotranspiration, infiltration, and channel routing, allowing for a comprehensive understanding of the hydrological system [13]. In addition, PBHMs can simulate river flows under various scenarios, enabling the assessment of water availability and allowing the design of water management strategies on river flow dynamics [14]. However, PBHMs require extensive data inputs, including accurate topographic information, soil properties, and meteorological data, which can be challenging to obtain. Model calibration can be time-consuming and sensitive to parameter estimation [15]. Additionally, these models may oversimplify certain processes or neglect important factors due to computational constraints.

Deep Learning (DL) models have been arising as an alternative modeling approach of streamflows in water bodies and reservoirs, offering several advantages in understanding and predicting water availability, enabling informed decision-making [16, 17, 18]. Similar to PBHMs, data-intelligence models have the potential to incorporate a wide range of variables, including climate indices, land use patterns, and reservoir operational rules, enabling a comprehensive understanding of the factors influencing streamflow dynamic [19]. Additionally, the ability to quantify prediction uncertainties and provide probabilistic forecasts can support risk-based decision-making in water management [20]. Consequently, the development and application of deep learning approaches for long-term streamflow predictions in water reservoirs have gained considerable attention among hydrologists and water resource specialists.

1.2 Research literature

In recent years, DL has emerged as a powerful technique for time series forecasting in hydrology [21, 22]. Its ability to automatically extract complex temporal patterns from large datasets makes it a promising tool for long-term streamflow prediction. The success of such methods is remarkable in short-term hydrological modeling, but the developments of approaches for long-term prediction are still evolving [23, 24]. One prominent deep learning architecture widely used is the Long Short-Term Memory (LSTM) network, designed to capture long-term dependencies in time series data by incorporating memory cells and gate mechanisms [25]. Recent studies have demonstrated the effectiveness of LSTM-based models in accurately predicting long-term streamflow with high temporal resolution [26]. For instance, Manavalan and Bynagari [27] proposed a multi-time scale LSTM framework for multiple timescale forecasting. Following this development, Cheng et al. [28] employed an LSTM network to forecast streamflow up to 20 days ahead, achieving superior performance compared to traditional hydrological models.

Another variant of DL architecture that has gained attention in long-term streamflow prediction is the Convolutional Neural Network (CNN) [29]. CNNs are particularly well-suited for capturing spatial and temporal features in hydrological data. Researchers have successfully applied CNN-based models to predict streamflow at different time scales, such as monthly or seasonal predictions [30]. In addition to LSTM and CNN architectures, hybrid models combining different deep learning components have also been explored. These models aim to leverage the strengths of multiple DL techniques to enhance prediction accuracy. For instance, Dehghani et al. [31] proposed a hybrid model that integrates LSTM and CNN layers, achieving improved performance in long-term streamflow prediction compared to individual models. Hybrid models are still relatively unexplored for long-term streamflow prediction, making it a promising area for developing artificial intelligence-based approaches.

Despite the recent research on DL in hydrological modeling, the full extent of its suitability for long-term streamflow prediction is still being explored. Huang, Qian, and Ochoa [32] proposed a framework that combined LSTM network and time series analysis techniques to address the scarcity of water temperature data for assessing thermal regimes. By reconstructing the monthly water temperature series from 1960 to 2020, reliable surrogate data was generated for thermal regime evaluation. To capture interdependencies within rainfall-runoff series, Chen et al. [33] combined the self-attention mechanism with a multi-layer LSTM model. Similarly, Noor et al. [34] incorporated a spatial attention layer preceding the LSTM layer and used chained equations for processing missing values. Wang et al. [35] introduced a novel spatiotemporal attention mechanism and developed an interpretation technique to examine the attention layer weights, providing insights into water level prediction in the medium to long term.

In addition to the above-mentioned approaches, other deep learning architectures, and other recent DL-based models, such as Temporal Fusion Transformer (TFT) [36] have also been widely explored for long-term streamflow prediction. The Temporal Fusion Transformer (TFT) is a state-of-the-art deep learning architecture specifically designed for time series forecasting. It combines the strengths of transformer models and LSTM networks, allowing for efficient encoding of temporal dynamics and capturing global dependencies [37]. TFT incorporates gating mechanisms, variable selection networks, static covariate encoders, and temporal processing layers to model complex hydrological systems effectively [35]. TFT has shown the ability to handle heterogeneous data sources, integrate static covariates, and capture long-term dependencies for long-term streamflow prediction in diverse hydrological settings. Wang and Tang [38] employed a TFT-based model for multi-step ahead streamflow forecasting, achieving superior performance compared to traditional LSTM and CNN models.

1.3 Research significance and motivation

Despite the increasing popularity of deep learning (DL) in hydrological modeling, the extent of its applicability for long-term streamflow prediction is still evolving and requires further investigation. While a substantial body of research has been dedicated to exploring various aspects of short-term streamflow prediction using DL models, there remains a scarcity of studies specifically focused on long-term streamflow prediction. This research gap necessitates a deeper understanding of the potential of DL models in accurately forecasting streamflow over extended time horizons. By addressing this gap in the literature, we can enhance our knowledge and expand the repertoire of tools available for long-term streamflow prediction, thereby enabling more effective water resources management and decision-making.

There are two categories of deep learning methods for multi-horizon forecasting: iterated approaches and direct methods [39]. Iterated approaches utilize one-step-ahead prediction models and recursively feed predictions into future inputs to obtain multi-step predictions. These approaches often employ Long Short-Term Memory (LSTM) networks, such as Deep AR and Deep State-Space Models. Transformer-based architectures have also been explored to enhance forecasting performance. Direct methods, on the other hand, generate forecasts for multiple predefined horizons at each time step. They employ sequence-to-sequence models, with LSTM or convolutional encoders summarizing past inputs and various techniques generating future predictions. This paper focuses on the iterated methods under-explored recently for hydrological modeling.

At the convergence of hydrology and machine learning, the motivation behind the Temporal Fusion Transformer deep learning model arises from the urgent need for accurate and robust long-term streamflow predictions, particularly within the intricate framework of complex hydrological systems. This becomes especially pertinent in the Funil reservoir in southeastern Brazil, where the water has multiple uses, ranging from agriculture and industrial to human consumption. In addition, the reservoir is located on a river that provides various environmental services. With the innovative use of the Temporal Fusion Transformer model in this context, we aim to fill part of the gap between traditional forecasting methods and the evolving challenges of climate change, ensuring informed decision-making and sustainable resource management and enhancing resilience in the face of changing hydrological patterns.

1.4 Research objectives

This paper introduces a novel study that employs a Temporal Fusion Transformer (TFT) model as a surrogate model to simulate the streamflow upstream of the Funil reservoir in southeast Brazil. The utilization of TFT as a surrogate model for streamflow modeling contributes to assessing the potential of the recently developed DL approaches in hydrological studies. This research contributes to advancing the understanding and application of TFT models in streamflow prediction and highlights their significance in optimizing water resource management practices.

This paper is organized as follows. Section 2 describes the Funil Reservoir and the multi-horizon forecasting approach using the Temporal Fusion Transformer (TFT) model. It explains the TFT model formulation and the time series prediction framework. Section 3 presents the computational experiments and evaluation of the proposed methodology using the Funil Reservoir dataset, comparing the TFT model's performance with baseline models while section 4 and discuss its strengths, limitations and potential for water resource management. The conclusion in Section 5 summarizes the findings and emphasizes the significance of the TFT model in accurately predicting long-term streamflow.

2. Material and Methods

2.1 Funil Reservoir

The Funil Reservoir in the Paraíba do Sul River is a vital resource with multiple water uses, ensuring water security for communities and agricultural productivity while also serving as a main source for industrial growth and economic development. Simultaneously, the reservoir contributes to environmental preservation, supporting biodiversity conservation and mitigating flooding risks in the downstream region.

The Funil Hydroelectric Power Plant dam is located in the middle course of the Paraíba do Sul River, in the municipality of Resende, in the Atlantic Plateau region (Figure 1), located at 22°35'S and 44°35'W. The reservoir is situated in the Southeast Coastal Basin, a region with intense chemical weathering, mountainous terrain, and a hot and rainy tropical climate in the summer and dry in the winter. With an area of 40 km^2 and a perimeter of 320 km, the Funil reservoir has rugged topography due to its location. Its average depth is 20 meters, and the water residence time is approximately 55 days [40]. The reservoir has been operating for over half a century, starting in 1969.

The Funil reservoir, particularly considering its location, presents a dense urban concentration, highly industrialized, which receives a large part of the domestic and industrial effluents produced and discharged upstream Currently, the Funil HPP lake is considered eutrophic, with an area of flooded Atlantic Forest and multiple land uses in its surroundings [41, 42]. Several studies indicate

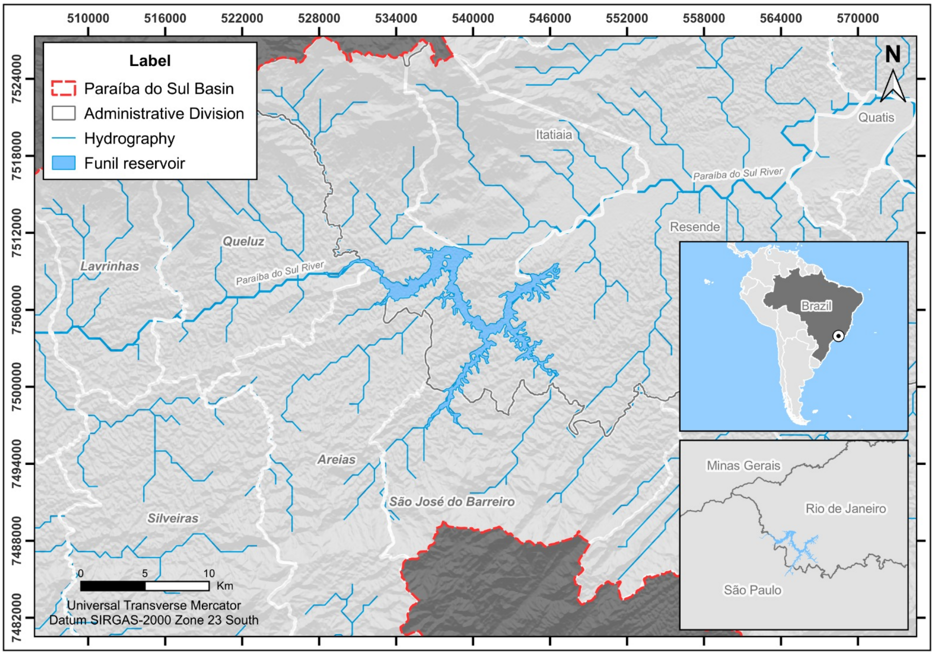

Figure 1: *Funil dam area.*

that the river is in an advanced state of eutrophication, with frequent occurrences of algal blooms, mainly of the genus *Microcystis aeruginosa* [43].

Figure 2 shows the geomorphological condition of the Funil dam area, including the topography, slope gradients, drainage patterns, and geological formations present in the Funil dam area.

Figure 3 shows the historical streamflow series of Funil reservoir inflow. The training data is in blue, and the test data is in green. Figure 4 depicts the Autocorrelation function (ACF) analysis conducted on the upstream river flow data of Funil reservoir. The ACF plot provides a visual representation of the correlation structure within the time series. As observed, the ACF values gradually decrease as the time lag increases, indicating the present observation maintains a correlation with its neighboring past observations; however, the strength of this correlation diminishes as the time lag between observations increases.

2.2 Multi-horizon forecasting

Multi-horizon forecasting involves predicting variables of interest at multiple future time steps, which is a critical problem in time series deep learning. Unlike one-step-ahead predictions[44, 45], multi-horizon forecasts provide users with estimates along the entire forecast path, enabling them to optimize actions at multiple future steps [36]. This has significant real-world applications where improved forecasting methods can yield valuable performance improvements. Figure 5 illustrates data sources for multi-horizon forecasting analyzed in this paper.

In various scenarios, prediction intervals are vital in optimizing decisions and managing risks by estimating the best and worst-case values the target variable can assume. To address this, we adopt quantile regression in a multi-horizon forecasting setting, generating forecasts for percentiles such as the 10th, 50th, and 90th at each time step. Each quantile forecast is defined as follows [36]:

$$\hat{y}_{\tau,t,\Delta} = F_{\tau}^{-1}\left(\mathbf{x}, \mathbf{u}_t; \theta\right) \quad (1)$$

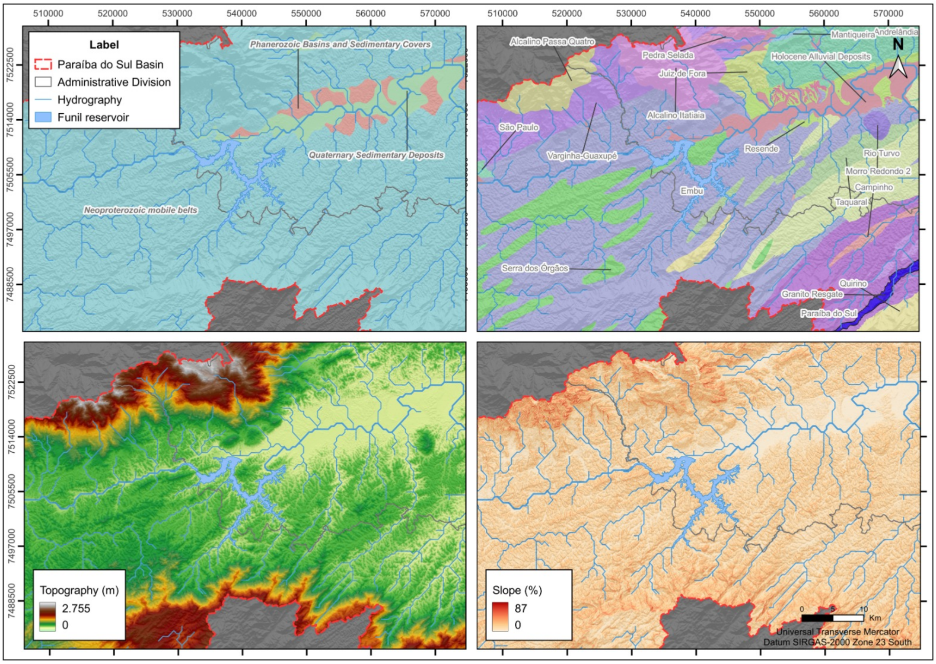

Figure 2: *Funil Reservoir: Hydrography,Topography, Geology, Slope.*

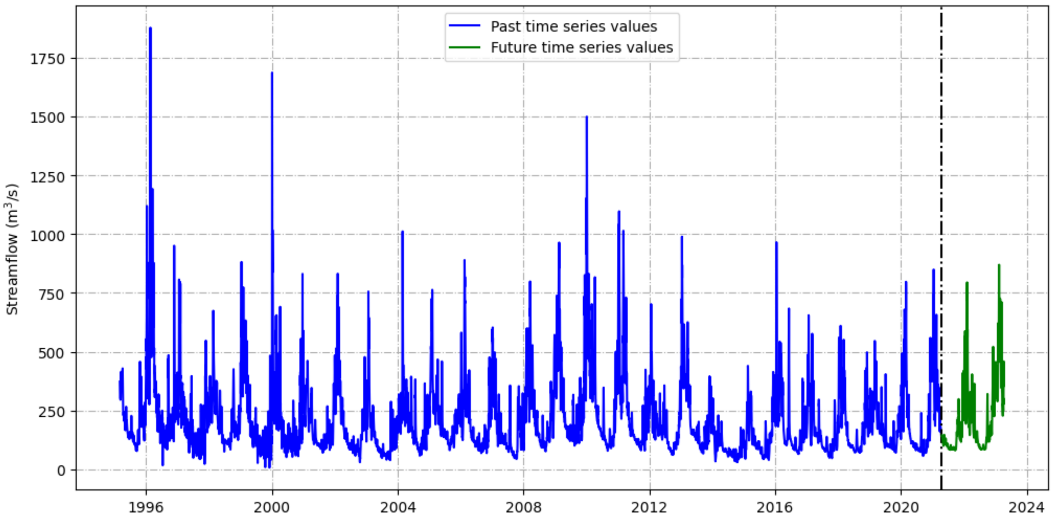

Figure 3: *Historical streamflow series of Funil reservoir inflow. The training data is shown in blue, and the test data is shown in green.*

where, $\hat{y}_{\tau,t,\Delta}$ represents the predicted τth sample quantile for the forecast Δ steps ahead at time t. The function $F_{\tau}^{-1}(\cdot)$ corresponds to the chosen prediction model. Following the approach of other direct methods, we generate forecasts simultaneously for Δ time steps, denoted as $\hat{y}_{\tau,t,\Delta} = [\hat{y}_{\tau,t+1,\Delta}, \hat{y}_{\tau,t+2,\Delta}, \ldots, \hat{y}_{\tau,t+\Delta,\Delta}]$. To incorporate past information, we utilize a finite look-back win-

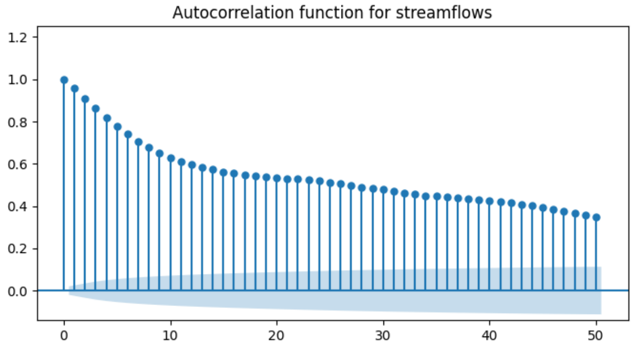

Figure 4: *Autocorrelation function (ACF) for upstream river flow of Funil reservoir.*

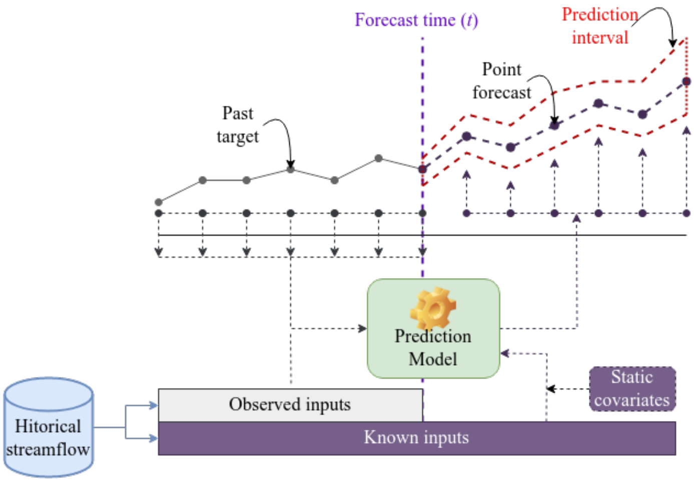

Figure 5: *Multi-horizon forecasting scenario. Adapted from [36].*

dow of size L, considering target and known inputs up to and including the forecast start time t. Specifically, $\mathbf{x}$ encompasses $[\mathbf{x}_{t-L}, \mathbf{x}_{t-L+1}, \ldots, \mathbf{x}_t]$, while $\mathbf{u}_t$ encompasses $[\mathbf{u}_{t-L}, \mathbf{u}_{t-L+1}, \ldots, \mathbf{u}_t]$. Furthermore, known inputs are considered across the entire range, denoted as $\mathbf{u}_k = [\mathbf{u}_{t-L}, \mathbf{u}_{t-L+1}, \ldots, \mathbf{u}_T]$.

2.3 Temporal Fusion Transformer (TFT)

Lim et al. [36] present the TFT architecture, which incorporates gated residual network blocks, LSTM for local processing, and multi-head attention for integrating information, demonstrating the effectiveness of TFT through use cases, showcasing its ability to interpret global behaviors and identify temporal dynamics in complex time series data. The mathematical formulation of the Temporal Fusion Transformer (TFT) method can be represented as follows.

Given a time series dataset with inputs $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T\}$ and targets $\mathbf{y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T\}$, where $\mathbf{x}_t$ represents the input features at time t and $\mathbf{y}_t$ represents the corresponding target values, TFT aims to learn a function

$$\mathbf{F} : \mathbf{X} \rightarrow \mathbf{y}$$

that maps the inputs to the targets. The TFT model consists of multiple components, including gating mechanisms, variable selection networks, static covariate encoders, and temporal processing modules. The gating mechanisms adaptively skip unused components, while the variable selection networks select relevant input variables at each time step. The static covariate encoders integrate static features into the model by encoding context vectors, and the temporal processing modules capture both long-term and short-term temporal relationships.

The mathematical equations of the TFT method can be written as follows:

1. Gating Mechanisms:
$$\mathbf{G} = \sigma(\mathbf{W}_{\text{gate}} \cdot \mathbf{x}) \tag{2}$$
2. Variable Selection Networks:
$$\mathbf{V} = \text{softmax}(\mathbf{W}_{\text{var}} \cdot \mathbf{G}) \tag{3}$$
3. Static Covariate Encoders:
$$\mathbf{C}_{\text{static}} = \text{Encoder}_{\text{static}}(\mathbf{x}_{\text{static}}) \tag{4}$$
4. Temporal Processing Modules:
$$\mathbf{C}_{\text{temporal}} = \text{Encoder}_{\text{temporal}}(\mathbf{x}_{\text{temporal}}) \tag{5}$$
5. Fusion Mechanism:
$$\hat{\mathbf{y}} = \text{Decoder}(\mathbf{C}_{\text{static}}, \mathbf{C}_{\text{temporal}}, \mathbf{V}) \tag{6}$$

where $\mathbf{G}$ represents the gating values, $\mathbf{V}$ represents the variable selection probabilities, $\mathbf{C}_{\text{static}}$ and $\mathbf{C}_{\text{temporal}}$ denote the static and temporal encodings, respectively, and $\hat{\mathbf{y}}$ represents the predicted target values.

The parameters $\mathbf{W}_{\text{gate}}$, $\mathbf{W}_{\text{var}}$, and the encoder/decoder weights are learned through optimization using a suitable loss function to predict the target values based on the input features accurately. Combining these equations allows the TFT model to effectively capture the temporal dependencies and accurately predict time series forecasting tasks.

The forecasting capability of TFT is further enhanced by its ability to provide prediction intervals. Through quantile forecasts, TFT estimates the range of likely target values at each prediction horizon, enabling users to assess uncertainty and make informed decisions. The following equation can represent the mathematical model of TFT:

$$\hat{y}_{\tau,t,\Delta} = F_{\tau}^{-1}(\mathbf{x}, \mathbf{u}_t; \theta) \tag{7}$$

where $\hat{y}_{\tau,t,\Delta}$ is the predicted τth sample quantile of the Δ-step-ahead forecast at time t, and $F_{\tau}^{-1}(\cdot)$ denotes the prediction model.

Figure 6 shows the high-level architecture of TFT, and its individual components.

2.4 Time Series Prediction Framework

Deep learning techniques have demonstrated remarkable efficacy when applied to time series prediction tasks, outperforming traditional statistical methods. These advanced methods leverage the power of complex neural networks to capture intricate temporal patterns and relationships within data providing accurate and insightful forecasts for future events.

The framework shown in Figure 7 is a general overview of the proposed deep learning approach to time series prediction. The framework consists of the following steps:

1. Data collection: The first step is to prepare the data for training the deep learning model from automatic stations in the river.
2. Covariates collection: The next step is to collect the information of interest and process them to be used as future covariates.
3. Model training and optimization: The third step is to train the model on the prepared data. This is done by using an optimization algorithm to minimize the loss function of the model.

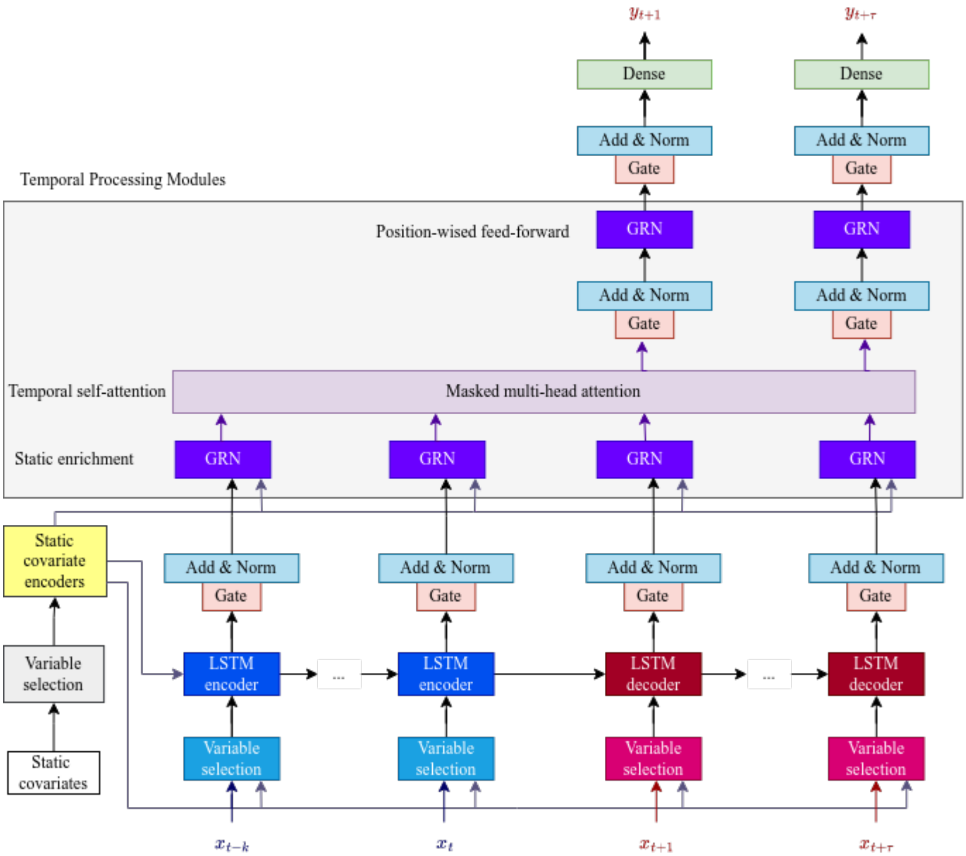

Figure 6: *The Temporal Fusion Transformer (TFT) model incorporates various inputs, including static covariates, time-varying past inputs, and a priori known future inputs. Variable Selection is employed to choose the most relevant features from the input intelligently. Gated information is introduced as a residual input, which undergoes normalization. Gated residual network (GRN) blocks optimize information flow by including skip connections and gating layers. Time-dependent processing involves using LSTMs for local processing alongside multi-head attention for information integration across different time steps.*

4. Long-term prediction: The final step is to deploy the trained model and employ the framework to generate long-term predictions.

This approach creates an implementation that results in superior performance, which can become a valuable tool for long-term flow forecasting, especially in complex hydrological systems.

3. Computational Experiments, Results

The computational experiments were conducted to evaluate the performance of the proposed methodology on the dataset collected from the Funil Reservoir. The dataset was partitioned into training and testing sets, as depicted in Figure 3, with the preservation of the temporal data order being ensured. The proposed methodology was implemented in Python programming language, using the autogluon package [46].

The TFT model can integrate covariates, which leverage the inherent patterns and correlations in the historical dataset to generate more precise and resilient predictions for the target variable. The

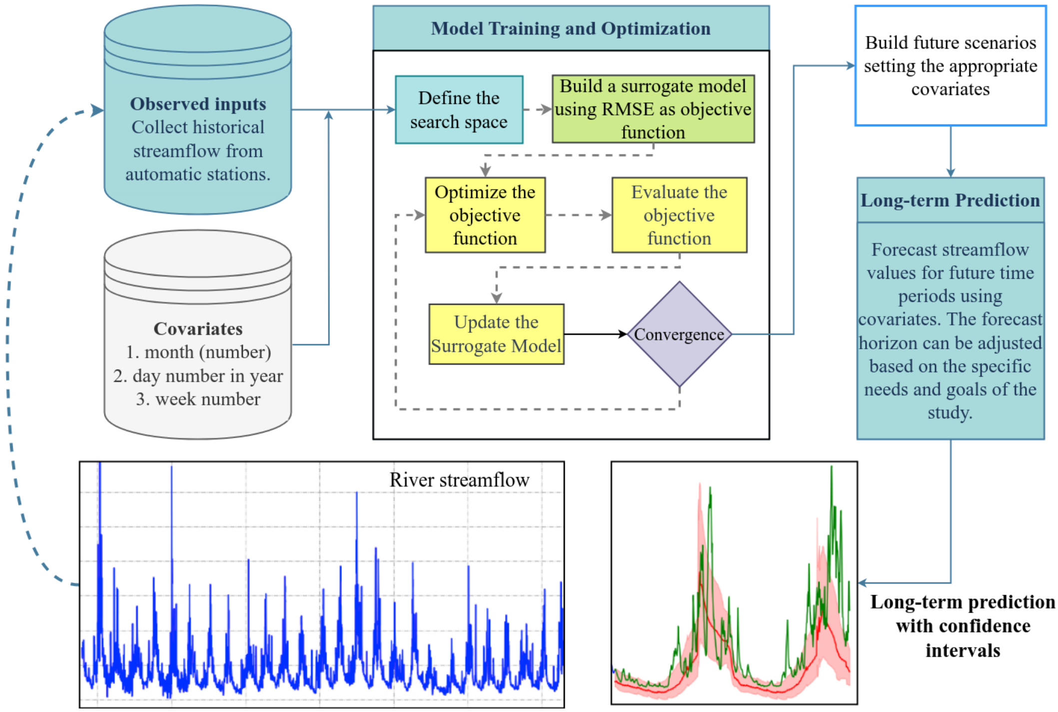

Figure 7: *Proposed long-term streamflow prediction framework using covariates.*

model implemented here considers the day of the year, the week of the year, and the month of the year, which provide temporal context to the model. Additionally, using past river flow as a historical covariate enhances the forecasting process by incorporating previous flow data to inform and refine predictions.

Table 1 summarizes the hyperparameters used in the Temporal Fusion Transformer (TFT) model within the computational framework, showing the details of the TFT model's configuration, including parameters that influence its behavior and performance. Each table row presents a specific hyperparameter, its corresponding description explaining its role within the model, and its default value.

Figure 8 compares the flow in the test set with the observed streamflow values. The hydrogram simulated by the TFT model showed good agreement with the observed hydrogram, indicating the model's ability to represent the complex dynamics of flow behavior. Despite not capturing in detail the variations in the observed river flow, the comparison validates the effectiveness of the TFT model in describing the hydrological processes of the study area. In addition, the predicted discharge confidence interval closely aligns with the observed discharge variation, providing additional assurance on the reliability of model predictions. It is important to note that these results were obtained only with historical flow data, highlighting the data-based nature of the TFT model.

Table 2 shows the results of the experiments that demonstrated the effectiveness of the proposed approach for long-term streamflow forecasting. The performance of our model was compared against several baseline models, including traditional statistical methods and deep learning techniques. The evaluation metrics employed comprised mean absolute error (MAE), root mean squared error (RMSE), Nash-Sutcliffe Efficiency (NSE), and coefficient of correlation (R).

The observation of the comparative table showing the different methods, including Seasonal Naive model [47], AutoARIMA [48], Theta method [49], Deep Arima model (DeepAR) [50],

Table 1: *TFT model hyperparameters.*

Parameter	Description	Value
context_length	Number of past time steps that are fed to the encoder. This hyperparameter controls the amount of historical data used to train the model.	64
hidden_dim	Hidden state size of the TFT, which controls the memory size the model uses to store the information it has learned about the data, and larger values more computational resources.	32
num_encoder_layers	Number of encoder layers that set the number of times that the encoder will process the data	3
num_decoder_layers	Number of decoder layers. It controls the number of times that the decoder will process the data.	32
num_heads	Number of attention heads in the self-attention layer in the decoder	4
dropout_rate	Dropout rate, used to regularize the model. A higher dropout rate will result in a more regularized model, which may help to prevent overfitting.	0.1
learning_rate	Learning rate. A smaller learning rate will result in a more gradual learning process, while a larger one will result in a faster learning process. The AdamW optimizer is a popular choice for deep learning models.	0.001
epochs	Number of epochs.	100
batch_size	Batch size is the number of data points used to update the model's parameters during each training iteration.	64
metrics	The metric to evaluate the model.	RMSE

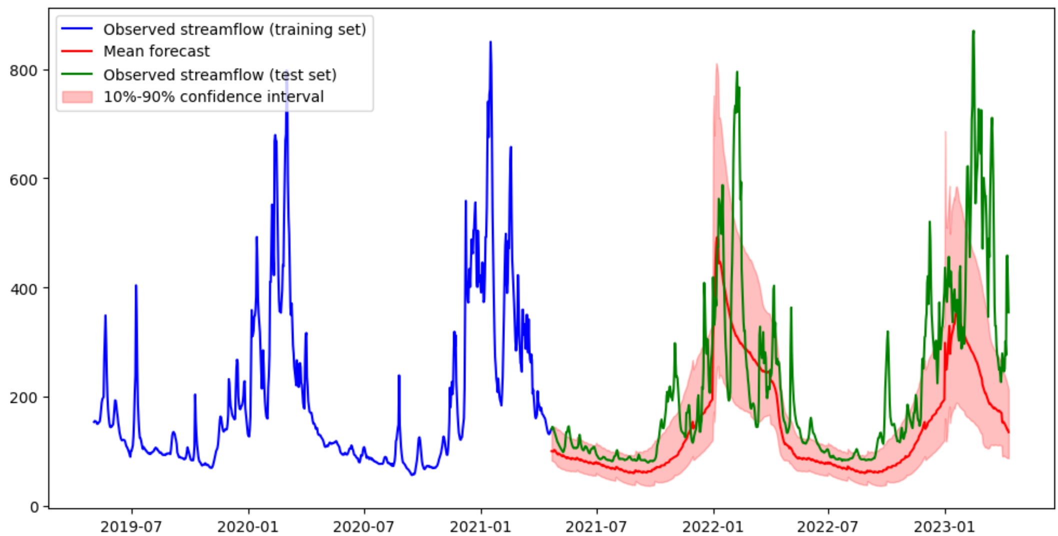

Figure 8: *Long-term streamflow prediction inflow Funil reservoir using the TFT model. The forecast is shown in red, the training data in blue, and the test data in green.*

and Temporal Fusion Transformer (TFT), reveals that the TFT consistently outperforms all other models in terms of streamflow prediction at a long-term scale. DeepAR, another DL-based model, demonstrates its strong predictive capabilities but slightly worse performance. Theta performs better than Seasonal Naive and AutoARIMA, indicating its effectiveness in capturing underlying patterns and trends in the data. As seen in the results reported in Table 2, TFT emerges as a highly promising and reliable model for long-term streamflow prediction, suggesting its potential applicability in real-world scenarios.

Table 2: *Comparison of approaches for long-term streamflow forecasting.*

	Model	MAE (m^3/s)	RMSE (m^3/s)	NSE	R
1	SeasonalNaive	113.35	181.12	-0.26	-0.00
6	AutoARIMA	113.98	168.67	-0.09	0.03
2	Theta	110.83	174.72	-0.17	0.45
4	DeepAR	77.84	137.59	0.27	0.72
5	TemporalFusionTransformer (TFT)	**70.88**	**121.66**	**0.43**	**0.79**

The results showed that the TFT model consistently outperformed the baseline and DL-based models. It exhibited superior accuracy, capturing both the short-term and long-term patterns in the streamflow data.

The TFT model uses a combination of attention mechanisms and gated residual networks to learn the long-term dependencies in streamflow data. It allows the TFT to effectively handle the temporal dependencies in the time series data, allowing it to make accurate forecasts even for long-term predictions.

The autoregressive approach implemented in the TFT model facilitates capturing the temporal dynamics in the time series data, allowing the incorporation of the previously predicted values as inputs for future predictions. The autoregressive mechanism enhances the accuracy and robustness of the TFT framework by considering the sequential nature of the data.

4. Discussion

The computational experiments confirmed the efficacy of the proposed methodology for long-term streamflow forecasting in the Funil Reservoir. The results demonstrate its potential for improving water resource management and decision-making in similar hydrological contexts. The findings also open avenues for future research, such as exploring the applicability of the proposed methodology in other hydrological regions and integrating additional data sources for enhanced forecasting accuracy.

Deep-learning and machine learning models offer distinct advantages compared to physically-based simulation models, such as SWAT and HMS-HEC. The main advantage is their independence from detailed knowledge of underlying physical processes and parameters, which can be challenging to obtain in complex and poorly gauged watersheds. By relying on patterns and relationships found in historical data, data-driven models can capture complex interactions not explicitly represented in physical models.

Furthermore, such models are computationally efficient and require fewer computational resources than their physically-based counterparts, with simpler structures and no need for solving complex systems of equations or time-consuming simulations. Despite the intricate mathematical formulation, data-driven models enable faster training and prediction, making them suitable for real-time or operational forecasting applications.

However, DL and ML models heavily depend on the quality and representativeness of available data, with limited or biased data potentially leading to inaccurate predictions and limited generalizability. Additionally, data-driven models may struggle to simulate extreme events for future scenarios.

In contrast, physically-based simulation models like SWAT and HMS-HEC explicitly represent physical processes and can incorporate detailed information on watershed characteristics, land use, and climate inputs. These models provide a more mechanistic understanding of the system and are better suited for analyzing land use impacts, evaluating management scenarios, and studying long-term hydrological responses. Integrating both approaches can offer complementary insights and enhance overall hydrological understanding and prediction capabilities.

It is important to note that AI-based predictions are built on a historical information base where the model aims to replicate the river flow dynamics. To a certain extent, these models learn about the hydrological dynamics and allow replicating the behavior for future steps. Although long-term AI-based models work with covariant information, which allows simulating scenarios, this information may not be sufficient to represent the complexity and non-linear relationships between the observed river flow and the hydrogeomorphological parameters. In such cases, what serves as an advantage of AI models, by not requiring a large mass of climate, geology, land use, and occupation information, can become the main drawback of using these models for long-term predictions.

In these scenarios, it is recommended that AI-based prediction models not replace physical-based models but are used to promote better decision-making for water resource management specialists who design water use policies or any other stakeholders involved in the processes associated with this important natural resource. For example, some physical-based models may face difficulties in modeling very large river basins, in the order of 30,000–100,000 km^2, due to the amount of information required or even the computational processing. In these cases, AI-based models, which work well with large datasets, can be used as an alternative to understand some dynamics and guide the parameterization of physical-based models. Furthermore, the reliability and forecasting capacity of AI-based models can vary depending on the characteristics of the river basins, and this process also involves a learning curve for users of these models.

The TFT model can effectively learn the underlying relationships between the historical streamflow data by leveraging deep learning techniques and attention mechanisms. Future research includes other covariates data such as rainfall, temperature, evaporation, and catchment attributes. This comprehensive approach allows the model to provide reliable predictions of the streamflow upstream of the reservoir, even in the face of increasing variability and uncertainty in hydrological processes.

Despite the extensive development of deep learning models for time-series forecasting, certain limitations persist within the context of hydrology. One such limitation is that deep neural networks often rely on time series data that is discretized at regular intervals [51], posing challenges when dealing with historical streamflow series that may contain missing observations or irregular arrival intervals [52]. Future research should address these challenges to enhance further the accuracy and reliability of deep learning models for long-term streamflow prediction. Further research involves feature selection [53] and integrating meta-heuristics with deep learning models, which has emerged as a promising strategy in recent research [54].

The results of this study demonstrate the efficacy of the TFT model in accurately predicting streamflow upstream of the Funil reservoir. The simulations conducted using the TFT model exhibit a high level of agreement with the observed streamflow data, successfully capturing seasonal patterns and long-term trends. This capability has the potential to empower water managers and decision-makers to make informed choices regarding reservoir operations and management. The research suggests that the TFT model is a promising new approach for streamflow prediction. The model is able to capture complex spatiotemporal patterns in the streamflow data, and it performs well in terms of accuracy. This makes the TFT model a potential tool for water managers and decision-makers who must control the reservoir operations and manage the water balance. However, it is important to note that the TFT model is still under development. Future research should focus on improving the model's accuracy and expanding its capabilities. Additionally, more research is needed to understand how the TFT model can be used to make informed decisions about reservoir operations and management.

This work contributes to developing and implementing models that address hydrological phenomena of surface hydrology in order to forecast streamflows in river basins. More specifically, to assess the water balance in the region upstream of Funil reservoir in the Paraiba do Sul river basin, covering urban, agricultural, and forested areas. The results of this research are of direct interest to public inspection, water usage authorization, and environmental control agencies, including the regional and national levels. Furthermore, this research can be of interest to the population living nearby the reservoir, the agricultural sector, industries, and hydroelectric power plants for power generation, whose risk of water shortages can be minimized by developing better hydrological forecasting models. In particular, the Funil reservoir will benefit greatly from the models developed in this paper.

5. Conclusion

This paper presents a study where a Temporal Fusion Transformer (TFT) model is utilized as a surrogate model to simulate the streamflow upstream of the Funil reservoir. The TFT model captures the streamflow data's complex spatiotemporal patterns and dependencies. The study's results demonstrate the effectiveness of the TFT model in accurately predicting the streamflow upstream of the reservoir. The study highlights the TFT model's potential as an effective surrogate model for simulating streamflow upstream of water reservoirs. By capturing complex spatiotemporal patterns and providing accurate predictions, the TFT model arises as a valuable tool for streamflow prediction and can potentially be helpful for water resource management and decision-making in reservoir operations. The current study emphasizes the importance of developing advanced prediction models to address the challenges of evolving hydrological conditions and facilitate sustainable water management practices. The findings presented in this paper highlight the potential of TFT models as valuable tools in water resource management. The accurate streamflow predictions obtained from the TFT model can contribute to optimizing reservoir operations, water allocation, and infrastructure management. Furthermore, the success of the TFT model in this study encourages further exploration and application of this approach in other water reservoir systems.

Acknowledgments

This work has been supported by ArcelorMittal do Brasil and UFJF's High-Speed Integrated Research Network (RePesq) https://www.repesq.ufjf.br.

Conflicts of Interest: **The authors have no conflict of interest to any part.**

References

[1] A. D. Martinho, C. M. Saporetti, and L. Goliatt. "Approaches for the short-term prediction of natural daily streamflows using hybrid machine learning enhanced with grey wolf optimization". In: *Hydrological Sciences Journal* 0.0 (2022), pp. 1–18. DOI: 10.1080/02626667.2022.2141121. eprint: https://doi.org/10.1080/02626667.2022.2141121. URL: https://doi.org/10.1080/02626667.2022.2141121.

[2] R. Rodriguez Núñez et al. "Water-quality data imputation with a high percentage of missing values: A machine learning approach". In: *Sustainability, vol. 13, no 11, pp. 1-17, jun 2021* (2021).

[3] R. M. A. Ikram et al. "Covariance Matrix Adaptation Evolution Strategy for Improving Machine Learning Approaches in Streamflow Prediction". In: *Mathematics* 10.16 (2022). ISSN: 2227-7390. DOI: 10.3390/math10162971. URL: https://www.mdpi.com/2227-7390/10/16/2971.

[4] T. R. Lopes et al. "Hydrological modeling for the Piracicaba River basin to support water management and ecosystem services". In: *Journal of South American Earth Sciences* 103 (2020), p. 102752.

[5] M. Di Modugno et al. "Build-up/wash-off monitoring and assessment for sustainable management of first flush in an urban area". In: *Sustainability* 7.5 (2015), pp. 5050–5070.

[6] D. P. Souza et al. "Group method of data handling to forecast the daily water flow at the Cahora Bassa Dam". In: *Acta Geophysica* (2022), pp. 1–13.

[7] A. Gorgoglione, A. Gioia, and V. Iacobellis. "A framework for assessing modeling performance and effects of rainfall-catchment-drainage characteristics on nutrient urban runoff in poorly gauged watersheds". In: *Sustainability* 11.18 (2019), p. 4933.

[8] A. Donigian Jr, B. Bicknell, J. Imhoff, et al. "Hydrological Simulation Program-Fortran (HSPF)." In: *Computer models of watershed hydrology.* (1995), pp. 395–442.

[9] G. Leavesley, L. Stannard, et al. "The precipitation-runoff modeling system-PRMS." In: *Computer models of watershed hydrology.* (1995), pp. 281–310.

[10] W. J. Charley. *The hydrologic modeling system (HEC-HMS): Design and development issues*. 149. US Army Corps of Engineers, Hydrologic Engineering Center, 1995.

[11] E. M. Schneiderman et al. "Modeling the Hydrochemistry of the Cannonsville Watershed with Generalized Watershed Loading Functions (GWLF) 1". In: *JAWRA Journal of the American Water Resources Association* 38.5 (2002), pp. 1323–1347.

[12] T. Liu et al. "Long-term streamflow forecasting for the Cascade Reservoir System of Han River using SWAT with CFS output". English. In: *Hydrology Research* 50.2 (2019). cited By 4, pp. 655–671. ISSN: 19989563. DOI: 10.2166/nh.2018.114.

[13] T. R. Lopes et al. "Multicriterial Analysis for the Identification of Key Areas to Improve the Management in Piracicaba River Basin, Brazil". In: *Journal of Hydrologic Engineering* 27.1 (2022), p. 05021033.

[14] U. R. V. Aires et al. "Machine learning-based modeling of surface sediment concentration in Doce river basin". In: *Journal of Hydrology* 619 (2023), p. 129320.

[15] C. Chen et al. "A novel paradigm for integrating physics-based numerical and machine learning models: A case study of eco-hydrological model". In: *Environmental Modelling & Software* 163 (2023), p. 105669.

[16] S. Heddam et al. "Cyanobacteria blue-green algae prediction enhancement using hybrid machine learning–based gamma test variable selection and empirical wavelet transform". In: *Environmental Science and Pollution Research* (June 2022). ISSN: 1614-7499. DOI: 10.1007/s11356-022-21201-1. URL: https://doi.org/10.1007/s11356-022-21201-1.

[17] U. R. V. Aires et al. "Modeling of surface sediment concentration in the Doce River basin using satellite remote sensing". In: *Journal of Environmental Management* 323 (2022), p. 116207.

[18] F. Vilaseca et al. "Daily rainfall-runoff modeling at watershed scale: a comparison between physically-based and data-driven models". In: *Computational Science and Its Applications–ICCSA 2021: 21st International Conference, Cagliari, Italy, September 13–16, 2021, Proceedings, Part VII 21*. Springer. 2021, pp. 18–33.

[19] A. Gorgoglione et al. "Influence of land use/land cover on surface-water quality of Santa Lucia river, Uruguay". In: *Sustainability* 12.11 (2020), p. 4692.

[20] C. Russo et al. "Improving the sediment and nutrient first-flush prediction and ranking its influencing factors: An integrated machine-learning framework". In: *Journal of Hydrology* 616 (2023), p. 128842.

[21] Z. C. Herbert, Z. Asghar, and C. A. Oroza. "Long-term reservoir inflow forecasts: enhanced water supply and inflow volume accuracy using deep learning". In: *Journal of Hydrology* 601 (2021), p. 126676.

[22] I.-F. Kao et al. "Exploring a Long Short-Term Memory based Encoder-Decoder framework for multi-step-ahead flood forecasting". In: *Journal of Hydrology* 583 (2020), p. 124631.

[23] F. Ghobadi and D. Kang. "Improving long-term streamflow prediction in a poorly gauged basin using geo-spatiotemporal mesoscale data and attention-based deep learning: A comparative study". In: *Journal of Hydrology* 615 (2022), p. 128608.

[24] M. Najafzadeh and S. Anvari. "Long-lead streamflow forecasting using computational intelligence methods while considering uncertainty issue". In: *Environmental Science and Pollution Research* (2023), pp. 1–17.

[25] K. M. Hunt et al. "Using a long short-term memory (LSTM) neural network to boost river streamflow forecasts over the western United States". In: *Hydrology and Earth System Sciences* 26.21 (2022), pp. 5449–5472.

[26] L. Ni et al. "Streamflow and rainfall forecasting by two long short-term memory-based models". In: *Journal of Hydrology* 583 (2020), p. 124296.

[27] M. Manavalan and N. Bynagari. "A Single Long Short-Term Memory Network can Predict Rainfall-Runoff at Multiple Timescales". In: *International Journal of Reciprocal Symmetry and Physical Sciences* 2 (2015), pp. 1–7.

[28] M. Cheng et al. "Long lead-time daily and monthly streamflow forecasting using machine learning methods". In: *Journal of Hydrology* 590 (2020), p. 125376.

[29] Y. Liu et al. "Directed graph deep neural network for multi-step daily streamflow forecasting". In: *Journal of Hydrology* 607 (2022), p. 127515.

[30] X. Shu et al. "Monthly streamflow forecasting using convolutional neural network". In: *Water Resources Management* 35 (2021), pp. 5089–5104.

[31] A. Dehghani et al. "Comparative evaluation of LSTM, CNN, and ConvLSTM for hourly short-term streamflow forecasting using deep learning approaches". In: *Ecological Informatics* 75 (2023), p. 102119.

[32] F. Huang, B. Qian, and C. Ochoa. "Long-term river water temperature reconstruction and investigation: A case study of the Dongting Lake Basin, China". English. In: *Journal of Hydrology* 616 (2023). cited By 1. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2022.128857.

[33] X. Chen et al. “The importance of short lag-time in the runoff forecasting model based on long short-term memory”. In: *Journal of Hydrology* 589 (2020), p. 125359.

[34] F. Noor et al. “Water level forecasting using spatiotemporal attention-based long short-term memory network”. In: *Water* 14.4 (2022), p. 612.

[35] Y. Wang et al. “Medium-long-term prediction of water level based on an improved spatio-temporal attention mechanism for long short-term memory networks”. In: *Journal of Hydrology* 618 (2023), p. 129163.

[36] B. Lim et al. “Temporal fusion transformers for interpretable multi-horizon time series forecasting”. In: *International Journal of Forecasting* 37.4 (2021), pp. 1748–1764.

[37] A. Kaur et al. “Fusion of multivariate time series meteorological and static soil data for multistage crop yield prediction using multi-head self attention network”. In: *Expert Systems with Applications* 226 (2023), p. 120098.

[38] C. Wang and W. Tang. “Temporal Fusion Transformer-Gaussian Process for Multi-horizon River Level Prediction and Uncertainty Quantification”. In: *Journal of Circuits, Systems and Computers* (2023).

[39] A. Alsharef et al. “Review of ML and AutoML solutions to forecast time-series data”. In: *Archives of Computational Methods in Engineering* 29.7 (2022), pp. 5297–5311.

[40] C. W. Branco et al. “Limnological features of Funil Reservoir (RJ, Brazil) and indicator properties of rotifers and cladocerans of the zooplankton community”. In: *Lakes & Reservoirs: Research & Management* 7.2 (2002), pp. 87–92.

[41] J. C. Rocha. “Abordagem sistêmica na gestão dos recursos hídricos de reservatórios hidrelétricos: aplicação no reservatório da UHE de Funil–RJ”. PhD thesis. Universidade Federal do Rio de Janeiro, 2021.

[42] I. L. R. Walter. “Análise de riscos em usinas hidrelétricas por abordagens complementares: estudo prático na usina de Funil”. In: (2019).

[43] C. Bobeda. “Isolamento, purificação e caracterização de microcistinas (hepta-peptídeos tóxicos) encontrados em uma floração de cianobactérias ocorrida no reservatório do Funil (Resende, RJ)”. PhD thesis, 235 pp. PhD thesis. PPG Química de Produtos Naturais, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, 1993.

[44] A. D. Martinho et al. “Extreme learning machine with evolutionary parameter tuning applied to forecast the daily natural flow at Cahora Bassa dam, Mozambique”. In: *Bioinspired Optimization Methods and Their Applications: 9th International Conference, BIOMA 2020, Brussels, Belgium, November 19–20, 2020, Proceedings 9*. Springer. 2020, pp. 255–267.

[45] A. D. Martinho, T. L. Fonseca, and L. Goliatt. “Automated Extreme Learning Machine to Forecast the Monthly Flows: A Case Study at Zambezi River”. In: *Intelligent Systems Design and Applications: 20th International Conference on Intelligent Systems Design and Applications (ISDA 2020) held December 12-15, 2020*. Springer. 2021, pp. 1314–1324.

[46] N. Erickson et al. “Autogluon-tabular: Robust and accurate AutoML for structured data”. In: *arXiv preprint arXiv:2003.06505* (2020).

[47] Z. Pala and R. Atici. “Forecasting sunspot time series using deep learning methods”. In: *Solar Physics* 294.5 (2019), p. 50.

[48] N. Tran and D. A. Reed. “Automatic ARIMA time series modeling for adaptive I/O prefetching”. In: *IEEE Transactions on parallel and distributed systems* 15.4 (2004), pp. 362–377.

[49] D. D. Thomakos and K. Nikolopoulos. “Forecasting multivariate time series with the theta method”. In: *Journal of Forecasting* 34.3 (2015), pp. 220–229.

[50] D. Salinas et al. “DeepAR: Probabilistic forecasting with autoregressive recurrent networks”. In: *International Journal of Forecasting* 36.3 (2020), pp. 1181–1191.

[51] A. Martinho, C. Saporetti, and L. Goliatt. “Hybrid machine learning approaches enhanced with grey wolf optimization to short-term prediction of natural daily streamflows”. In: *Hydrological Sciences Journal* (2022). doi: 10.1080/02626667.2022.2141121.

[52] B. Lim and S. Zohren. “Time-series forecasting with deep learning: a survey”. In: *Philosophical Transactions of the Royal Society A* 379.2194 (2021), p. 20200209.

[53] L. Goliatt and Z. M. Yaseen. “Development of a hybrid computational intelligent model for daily global solar radiation prediction”. In: *Expert Systems with Applications* 212 (2023), p. 118295. issn: 0957-4174. doi: 10.1016/j.eswa.2022.118295. url: https://www.sciencedirect.com/science/article/pii/S0957417422014300.

[54] L. Goliatt et al. “Performance of evolutionary optimized machine learning for modeling total organic carbon in core samples of shale gas fields”. In: *Petroleum* (2023). issn: 2405-6561. doi: https://doi.org/10.1016/j.petlm.2023.05.005. url: https://www.sciencedirect.com/science/article/pii/S2405656123000354.