Energy 356 (2026) 141214

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

Optimized Harmonic-Multiscale Kolmogorov–Arnold Network denoised by EWT for hydroelectric dam useful volume forecast

Stefano Frizzo Stefenon [a,b,c,*], Laio Oriel Seman [e], João Pedro Matos-Carvalho [d], Ademir Nied [c], Gabriel Villarrubia Gonzalez [f], Kin-Choong Yow [b]

[a] Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa, Rua Conselheiro Emídio Navarro 1, 1959-007 Lisboa, Portugal
[b] Faculty of Engineering and Applied Sciences, University of Regina, Saskatchewan, Canada
[c] Universidade do Estado de Santa Catarina (Udesc), Centro de Ciências Exatas e Tecnológicas, Departamento de Engenharia Elétrica, Programa de Pós-Graduação em Engenharia Elétrica, Grupo de Pesquisa em Controle de Sistemas, Joinville - SC, Brasil
[d] LASIGE, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal
[e] Department of Automation and Systems Engineering, Federal University of Santa Catarina, Florianópolis, Brazil
[f] Expert Systems and Applications Lab, University of Salamanca, Salamanca, Spain

ARTICLE INFO

Dataset link: https://dados.ons.org.br/dataset/ear-diario-por-reservatorio

Keywords:
Hydroelectric power
Reservoir forecasting
Kolmogorov–Arnold Network
Time series analysis
Harmonic-multiscale modeling

ABSTRACT

Accurate forecasting of the useful volume of a hydroelectric reservoir is essential for energy planning and water resource management. To address this challenge, this paper proposes a novel Harmonic-Multiscale Kolmogorov–Arnold Network forecasting model. The model integrates sinusoidal (Sin) basis functions to model global oscillatory patterns and wavelet (Wav) bases to capture localized transient features within the Kolmogorov Arnold Network (KAN) framework (SinWavKAN). This hybrid representation enhances feature expressiveness by jointly modeling smooth long-term dependencies and short-term irregular dynamics. The input signals are first denoised using the Empirical Wavelet Transform (EWT) to isolate informative modes from noise, and the SinWavKAN's hyperparameters are optimized via an Adaptive Tree-structured Parzen Estimator (ATPE). An ablation study confirmed the synergistic superiority of the sinusoidal-wavelet combination over other basis functions. Compared to state-of-the-art KAN variants (e.g., fastKAN, PyKAN, WavKAN), the proposed architecture, named in short Opt-EWT-SinWavKAN, demonstrated a reduction in mean absolute error by approximately 85%–93% across forecasting horizons from 1 to 30 days. The results conclusively show that the Opt-EWT-SinWavKAN provides a more reliable approach for hydroelectric reservoir volume forecasting.

1. Introduction

Hydroelectric power accounts for 47.8% of Brazil's total energy generation capacity, which generated 225,386 MW in 2024, making it the backbone of the country's electrical power system [1]. The reliable operation of this extensive hydroelectric network depends on effective reservoir management strategies since reservoirs are, in many cases, interconnected. A key challenge in this domain is accurately predicting the useful volume of reservoirs, which directly impacts power generation capacity and system reliability [2].

The useful volume of a reservoir represents its operational water storage capacity and serves as an indicator of its potential for power generation [3], considering that poor resource management can result in energy crises [4]. Traditional approaches to reservoir volume prediction have primarily relied on time series analysis and statistical methods [5]. However, these methods often fail to capture the complex interdependencies between connected reservoirs, particularly in regions with dense hydroelectric networks, resulting in poor performance in forecasting [6].

From an operational perspective, improving forecasting performance over horizons of 1 to 30 days has direct implications for hydroelectric system management. More reliable short and medium-term reservoir volume predictions can support generation dispatch planning, allowing operators to better allocate water resources according to expected inflows and demand conditions [7]. Improved forecasts may also help reduce unnecessary spill events by enabling earlier operational adjustments, thereby increasing energy efficiency and minimizing water losses. In addition, more accurate predictions

* Corresponding author at: Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa, Rua Conselheiro Emídio Navarro 1, 1959-007 Lisboa, Portugal.
E-mail addresses: stefano.stefenon@isel.pt (S.F. Stefenon), laio@gos.ufsc.br (L.O. Seman), jpecarvalho@ciencias.ulisboa.pt (J.P. Matos-Carvalho), ademir.nied@udesc.br (A. Nied), gvg@usal.es (G.V. Gonzalez), Kin-Choong.Yow@uregina.ca (K.-C. Yow).

https://doi.org/10.1016/j.energy.2026.141214
Received 6 February 2026; Received in revised form 23 March 2026; Accepted 29 April 2026
Available online 4 May 2026

contribute to risk mitigation by supporting proactive decision-making in situations involving drought conditions or flood risks, improving both system reliability and safety [8].

Kolmogorov Arnold Network (KAN)-based models have recently gained significant attention in time series forecasting due to their strong expressive power and ability to model complex nonlinear relationships [9]. However, the growing popularity of KAN has also led to the emergence of numerous architectural variations [10], each emphasizing different functional bases, parameterizations, and training strategies, which makes their direct application to real-world problems non-trivial [11].

In the context of useful volume forecasting in hydroelectric power plants, no single KAN variant is universally optimal, since the dynamics involve seasonal patterns, long-term dependencies, and abrupt regime changes driven by hydrological and operational factors. An optimized hybrid KAN model, combining complementary representations within a unified framework, can better capture these heterogeneous dynamics, leading to improved accuracy and robustness compared to individual KAN formulations when applied to operational forecasting tasks in hydroelectric systems.

Despite recent advances in hydrological forecasting and the emergence of KAN variants, there remains a lack of hybrid architectures capable of simultaneously capturing global periodic behavior, localized transient dynamics, and noise robustness in long-horizon reservoir volume prediction. In particular, existing approaches typically address representation learning, denoising, and hyperparameter optimization as separate steps, which may limit their effectiveness when applied to complex, nonstationary hydrological time series. To address this gap, this work investigates the following research question: Can a unified harmonic multiscale KAN architecture, combined with empirical wavelet-based denoising and adaptive hyperparameter optimization, improve the accuracy and robustness of hydroelectric reservoir useful volume forecasting? To answer this question, we propose a hybrid framework designed to integrate spectral and multiresolution representations within the KAN paradigm, and we evaluate its effectiveness through benchmarking and ablation studies against state-of-the-art forecasting models.

The proposed Harmonic Multiscale Kolmogorov Arnold Network is a KAN architecture in which each univariate edge function is modeled through a joint harmonic and multiscale representation, integrating Sinusoidal spectral components with Wavelet bases; based on these functions, our model is named SinWavKAN. The harmonic branch captures smooth global oscillatory behavior through a flexible spectral expansion that generalizes classical Sinusoidal parameterizations, while the multiscale branch introduces localized basis functions with learnable scales and translations, enabling the modeling of nonstationary and transient dynamics across multiple resolutions.

By embedding both representations within the Kolmogorov–Arnold additive framework, SinWavKAN preserves theoretical consistency while substantially increasing expressive power, allowing the network to simultaneously represent long-term periodic structures and localized variations that are typical in complex time series such as hydrological signals. To further improve the model, we performed hypertuning by an Adaptive Tree-structured Parzen Estimator (ATPE), ensuring its best setup. Additionally, we considered the Empirical Wavelet Transform (EWT) for signal denoising. Based on that, our model, named Opt-EWT-SinWavKAN, makes the following key contributions:

- **Novel Hybrid KAN Architecture:** Proposes the Harmonic-Multiscale Kolmogorov–Arnold Network, a new model that integrates sinusoidal expansions and wavelet bases within a unified framework to simultaneously model global oscillatory patterns and localized transient dynamics in hydrological time series.
- **Integrated Denoising Preprocessing:** Incorporates the EWT as a dedicated signal processing step to filter high-frequency noise from reservoir data, enhancing the model's ability to learn underlying trends and improving forecast robustness.
- **Adaptive Hyperparameter Optimization:** Employs ATPE to automatically tune the critical hyperparameters of SinWavKAN (e.g., number of harmonics and wavelets), ensuring an optimal model configuration without manual trial-and-error.
- **Comprehensive Ablation Study:** Conducts an extensive ablation analysis to validate the design choices, demonstrating the superior synergy between sinusoidal and wavelet components over other basis functions (e.g., Fourier, Chebyshev, Radial Basis Function (RBF)) and establishing the rationale for the proposed hybrid approach.
- **Rigorous Experimental Benchmarking:** Provides an exhaustive evaluation against multiple state-of-the-art KAN models across various forecasting horizons (1–30 days) and input window sizes, showing consistent and substantial performance gains (e.g., MAE reductions of 85%–93%).
- **Demonstrated Practical Superiority:** Confirms that the proposed model achieves significantly slower error growth with increasing prediction horizon, highlighting its enhanced capability for long-term forecasting and its potential for real-world application in hydroelectric reservoir management and energy planning.

The combination of sinusoidal and wavelet bases is motivated by the physical characteristics of reservoir volume dynamics, which typically include both periodic and transient behaviors. Sinusoidal functions are well-suited to represent seasonal and long-term hydrological cycles, while wavelet bases are effective for modeling localized variations caused by abrupt inflow changes and operational interventions. By integrating these two representations, the proposed SinWavKAN provides a theoretically consistent multiscale modeling approach that reflects the nonstationary and heterogeneous nature of reservoir dynamics, rather than relying solely on empirical performance improvements.

The use of EWT in the proposed framework is motivated not only by its ability to reduce high-frequency noise but also by its capacity to enhance the multiscale structure of the input signal in a way that is consistent with the SinWavKAN architecture. By decomposing the reservoir time series into frequency bands and reconstructing the signal from the most informative modes, EWT allows the model to focus on the dominant temporal patterns while attenuating stochastic fluctuations that may degrade learning stability. This preprocessing step also reinforces the architectural design of SinWavKAN, since the denoised signal better exposes both the global oscillatory components captured by the sinusoidal basis and the localized variations modeled by the wavelet basis.

The novelty of the proposed approach lies in the structural formulation of a harmonic multiscale KAN architecture in which sinusoidal and wavelet bases are jointly embedded as complementary univariate function representations within the Kolmogorov–Arnold framework. While elements such as wavelet preprocessing, KAN variants, and ATPE-based hyperparameter optimization have been explored individually in previous studies, their unified formulation in a single interpretable architecture designed specifically for multiscale hydrological forecasting represents the main conceptual contribution of this work. In particular, the proposed SinWavKAN advances the KAN design by explicitly structuring the edge functions to simultaneously capture global spectral behavior and localized temporal variations, providing a more physically consistent representation of nonstationary reservoir dynamics.

Unlike FourierKAN, WavKAN, or SirenKAN, which rely on a single type of functional basis, the proposed approach introduces a dual representation in which global spectral functions and localized multiscale functions are jointly learned within the same edge mapping. This design changes the functional approximation strategy from a single basis expansion to a cooperative decomposition, where different function families are explicitly assigned complementary representational roles. As a result, the contribution is not only empirical, but structural, as it extends the KAN formulation toward multi-basis functional representations specifically designed for nonstationary time series, providing a clearer theoretical distinction from existing single-basis KAN variants.

The rest of this paper is organized as follows: Section 2 provides background about hydropower inflow analysis. Section 3 explains how the water balance works in hydroelectric power plants. Section 4 details our methodology for forecasting useful levels of dams. Section 5 presents experimental results and analysis. Finally, Section 6 concludes the paper and discusses future research directions. Considering the vast nomenclature used in the field, for standardization.

2. Related works

Hydropower systems, especially in interconnected cascaded reservoirs like those in Brazil, are increasingly vulnerable to climate variability, inflow uncertainty, and operational interdependencies. Traditional process-based hydrological models have long been used to quantify how precipitation seasonality and rising temperatures influence reservoir recharge and overall hydropower reliability [12]. While effective for broad sensitivity analyses, these approaches often struggle with the nonlinear and non-stationary nature of real-world hydroclimatic data.

The rise of artificial intelligence and Machine Learning (ML) has shifted the focus toward more flexible, data-driven forecasting. Recurrent neural networks and transformers, for example, have improved inflow predictions at hydroelectric plants by better capturing temporal dependencies linked to dam level fluctuations and energy planning needs [13]. Building on this, ensemble methods such as Random Forest (RF) have proven particularly effective for reservoir inflow and hydropower production simulation. Obahoundje et al. [14] demonstrated this by using lagged hydroclimatic predictors (precipitation and temperature), k-fold cross-validation, and iterative feature refinement, achieving Pearson correlations above 0.90 for inflow and 0.80 for energy generation on monthly datasets, highlighting RF's strength in modeling nonlinear relationships even with limited data.

Beyond short-term forecasting, research has explored the practical value of imperfect long-term inflow predictions in operational settings. Shu et al. [15] showed through simulations that even low-accuracy forecasts can deliver substantial benefits (annual generation gains of 14.86–29.58 million kWh) when operational rules tolerate errors or when forecast biases align consistently with actual trends, emphasizing the role of inherent system resilience.

To explicitly handle inflow uncertainty in large cascade systems, probabilistic frameworks have gained traction. Lei et al. [16] employed copula-based synthetic runoff generation and Gibbs sampling to model dependencies, then used a mean–variance approach to balance energy generation against ecological risk, revealing that modest energy sacrifices (e.g., 1% reduction) can yield significant ecological improvements (6.14% risk decrease). Similarly, Zhao et al. [17] integrated uncertainty into scenario-based multi-stage optimization via a multi-agent reinforcement learning progressive optimality algorithm, enabling robust policies that coordinate spatial storage and temporal trajectories more effectively than deterministic baselines, while maintaining computational feasibility.

Other studies have addressed specific operational challenges in cascades. Zhang et al. [18] developed a river-aware framework for daily-regulated plants, using peak-shaving control and Alpha shapes to derive practical release rules that mitigate upstream–downstream effects. Yang et al. [19] applied an enhanced Chimp optimization algorithm for long-term scheduling across diverse hydrological regimes, reducing spillage and boosting generation (9.83–20.12% increases when forecasts meet skill thresholds). Broader climate-related risks have also been examined: Huang et al. [20] quantified compound extremes (flood-to-drought transitions), linking them to generation losses and greater resilience in highly regulated plants; Zhou et al. [21] attributed uncertainty sources in cascades under climate change, showing dominant roles for hydrological models in flood seasons; and He et al. [22] proposed dynamic seasonal flood-limited levels to improve generation (3.16% increase) with minimal added flood risk.

The economic value of forecasts has been assessed as well. Liu et al. [23] used stochastic dynamic programming to show reliability gains of 33.89% from 10-day-ahead predictions, with Bayesian variants further enhancing robustness against bias. For computational scalability in large systems, Niu et al. [24] accelerated the Progressive Optimality Algorithm (POA) via Long Short-Term Memory (LSTM)-coupled response surfaces, cutting runtime by 79.2% in dry-runoff scenarios.

Hyperparameter tuning methods have gained increasing attention in scientific research due to their ability to significantly improve the performance and robustness of ML models. As modern predictive architectures become more complex and highly sensitive to parameter configurations, the selection of appropriate hyperparameters emerges as a step in achieving reliable and reproducible results [25]. Consequently, optimization strategies such as Bayesian optimization, evolutionary algorithms, and multi-agent optimization approaches are being widely adopted across different tasks. These methods enable automated exploration of the search space, reducing manual effort while improving model performance.

Recent advances in interpretable deep learning, particularly KANs, offer promising alternatives for hydrological time series by replacing fixed activations with learnable splines for better nonlinearity capture and transparency. Applications include streamflow forecasting where KANs outperform Transformers in short horizons [26], hybrid KAN-LSTM models for permafrost-affected rivers [27], and attention-enhanced variants for water level prediction [28]. Wavelet-based extensions (e.g., Wav-KAN) further improve multi-scale handling [29], while EWT denoising has boosted reservoir/streamflow models in hybrid setups [30]. Hyperparameter tuning via ATPE has also shown efficiency in complex ML tasks [31]. An overview of the principal contributions in the literature on forecasting hydropower inflows and reservoir volumes, as well as on the optimization of cascaded hydropower systems, is presented in Table 1.

Despite these developments, key gaps remain: Most approaches treat forecasting and optimization in isolation, with limited integration of advanced interpretable models (like hybrid KANs), explicit denoising (EWT), and adaptive tuning (ATPE) for long-horizon, non-stationary reservoir volume prediction under real operational uncertainty. This work addresses these by introducing Opt-EWT-SinWavKAN, a novel Harmonic-Multiscale KAN that unifies sinusoidal global patterns and wavelet transients, denoised via EWT and optimized with ATPE, to deliver superior accuracy and robustness for hydroelectric dam useful volume forecasting.

3. Problem definition

The rainfall–runoff simulation outcomes are fed into the chain for the weekly price of settlement of differences forecasting, which keeps the electrical system economically stable for the operator [13]. The Brazilian National System Operator (ONS) uses the Soil Moisture Accounting Procedure (SMAP) mathematical approach in their weekly natural inflow forecast for the monthly operation program [34]. Besides the economic aspect, the flooding issue is widely discussed, being a serious problem due to the inability of rainfall in southern Brazil to flow during rainy periods [35]. This makes flow control, which is directly related to the level of the dam, very important for energy planning and the safety of the dam and its surroundings [36].

3.1. SMAP model

The SMAP needs daily potential evapotranspiration rates (E_{p_t}), observed daily average flows (Q_{obs_t}), and forecasted and observed daily precipitation as inputs for its operating mode [37]. Using a weighted average of readings from individual rainfall stations, the

Table 1
Summary of key related works on hydropower inflow/reservoir volume forecasting and cascade optimization.

Reference	Approach/Model	Key Focus/Application	Main Results/Strengths	Limitations/Gaps
[12]	Process-based hydrological models	Inflow sensitivity to climate variability	Quantifies precipitation/temperature impacts on recharge/reliability	Limited handling of nonlinear/non-stationary dynamics
[13,30]	Recurrent Neural Networks, Transformers	Inflow forecasting for dam levels/energy planning	Improved temporal dependency capture	May struggle with extremes/long horizons without hybrids
[14]	RF k-fold cross validation with lagged inputs	Reservoir inflow & hydropower production	Pearson corr. >0.90 (inflow), >0.80 (energy); strong on monthly data	Relies on limited predictors; less effective for extremes/long leads
[15]	Operational simulations with imperfect forecasts	Value of long-term forecasts in cascades	Gains 14–30 GWh/year even with low accuracy; error tolerance via rules	Benefits conditional on bias alignment; higher accuracy yields more
[16]	Copula + Gibbs sampling + mean–variance risk	Probabilistic risk in large cascades	1% energy sacrifice → 6.14% ecological risk reduction	Computationally intensive for very large systems
[17]	Multi-agent + progressive optimality	Scenario-based robust multi-stage scheduling	Risk-robust policies; efficient action space compression	Focus on coordination; limited explicit denoising/interpretability
[18]	River-aware peak-shaving + Alpha shapes	Upstream–downstream effects in daily-regulated plants	Mitigates oscillations; derives practical release rules	Short-term focus; less on long-horizon uncertainty
[19]	Improved Chimp optimization method	Long-term cascade scheduling across regimes	Reduced spillage; 9.83–20.12% generation increase with good forecasts	Metaheuristic; may not scale to extreme uncertainty
[20,21]	Global simulations + uncertainty attribution	Compound extremes & climate change impacts	Losses 9%–11%; seasonal uncertainty breakdown (40% in floods)	Diagnostic; limited predictive integration
[22]	Dynamic seasonal flood-limited levels	Flood season policies in cascades	3.16% generation increase with marginal flood risk	Policy-focused; assumes good seasonal data
[23]	Stochastic dynamic programming	Economic value of 10-day forecasts	33.89% reliability increase; Bayesian improves robustness	Synthetic forecasts; bias sensitivity
[24]	Response surface-based POA + LSTM	Scalable multireservoir optimization	79.2% computation reduction	Acceleration focus; less on interpretability
[26,27,32]	Kolmogorov–Arnold Networks (KAN, hybrids, attention)	Streamflow/runoff/water level forecasting	Superior short-term performance; Nash–Sutcliffe Efficiency up to >0.95	Emerging in hydrology; limited long-horizon/multi-scale hybrids
[30]	Signal decomposition by EWT + reservoir computing	Multi-step streamflow in Brazilian plants	Reduced errors in forecasting floods via denoising	Decomposition-specific; not unified with advanced networks
[31]	Adaptive Tree-structured Parzen Estimator	Hyperparameter tuning in ML/time series	Computing efficient in high-dimensional spaces	General; few hydrology-specific applications
[33]	Mixture of experts (MoE) enhanced transformer	Hydroelectric reservoir volume prediction	EWT + HeadComposer + MoE improves feature extraction	Higher model complexity and computational cost

model computes the basin-wide average daily precipitation, $P_{b(t)}$, given by:

$$P_b(t) = P_1(t)ke_1 + P_2(t)ke_2 + \cdots + P_n(t)ke_n. \tag{1}$$

The average rainfall in the basin at time t is represented by $P_b(t)$, while the observed rainfall at every location is indicated by $P_1(t), P_2(t), \ldots, P_n(t)$. The spatial representation weights for each station are represented by the coefficients $ke_1, ke_2, \ldots, ke_n$, which ensure that their sum equals 1, according to:

$$ke_1 + ke_2 + ke_3 + \cdots + ke_n = 1. \tag{2}$$

The model then uses the following equation to determine the rainfall that is thought to be indicative of the day $t(P_d(t))$, which is made up of a weighting of $P_b(t)$ of various times:

$$P_d(t) = [P_b(t-n)k(t-n)+\cdots+P_b(t)k(t)+P_b(t+1)k(t+1)+P_b(t+2)k(t+2)] \tag{3}$$

where n are days used in the weighting procedure, and $k(t-n)$, $\ldots, k(t), k(t+1), k(t+2)$ are the temporal weights. To ensure the basin's water balance, the $P_d(t)$ is multiplied by the factor P_{cof} (rainfall compensation coefficient) [38]. This yields the average precipitation in the basin $Precp(t)$, which the model will take into account at that moment,

$$Precp(t) = P_d(t) \cdot P_{\text{cof}}. \tag{4}$$

3.2. Water balance

The deterministic physical model (SMAP) is used for water balance, which is based on dividing the flow into surface runoff and subsurface runoff. The model stores each kind of runoff in hypothetical reservoirs that specify the form and length of the flows [39]. The operation of the Brazilian electrical system has been planned in large part thanks to this model. With Q representing the flow between the reservoirs and the precipitation on time E_{t0}, Fig. 1 illustrates this method.

The part of surface runoff is moved to the surface reservoir (R_{sup}), where water transfer ($marg$) to the lowland reservoir (R_{sup2}) and surface runoff (E_d and E_{d3}) are computed. The last one serves as a buffer for drainage channels and includes a calculation of the lowland surface runoff (E_{d2}) and an evaporation component (E_{marg}). The part that penetrates the soil is the distinction between surface runoff and rain precipitation. Assuming the soil field capacity limit has not been reached, it is moved to the reservoir (R_{soil}), where the real evapotranspiration (E_r) part is extracted, and also transmitted to the underground reservoir (R_{sub}) through recharge (R_{ec}). Subsurface runoff (E_b), which is the amount of water that seeps into the soil and replenishes groundwater, is computed in the R_{sub} reservoir [25].

The useful volume of hydroelectric dams is crucial for research in many industries, especially the electrical industry and flood control. Planning the operation of the hydroelectric power plants of the interconnected electrical power grid is made possible by a reliable flow estimate and dam level [40]. Forecasting both short- and long-term generation and spillover is part of this planning. The interconnection between the reservoirs motivates the use of graph-based models since the reservoirs are connected.

3.3. Volume management

The ONS's operation programming division publishes the daily plant operation program, an electro-energetic program for power plant operation [41]. Along with the forecast of the conditions for the operation of the reservoirs, this program includes the additions or modifications of operational limitations of the national energy generation and transmission facilities, the forecast of daily load, and the programming of automatic generation control.

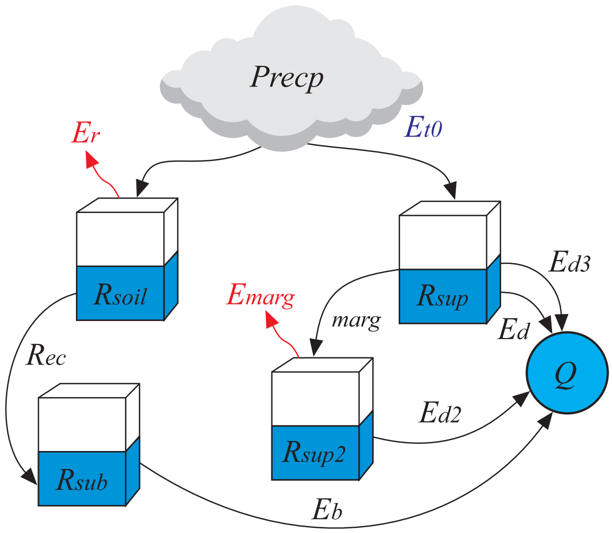

Fig. 1. Water balance used model.

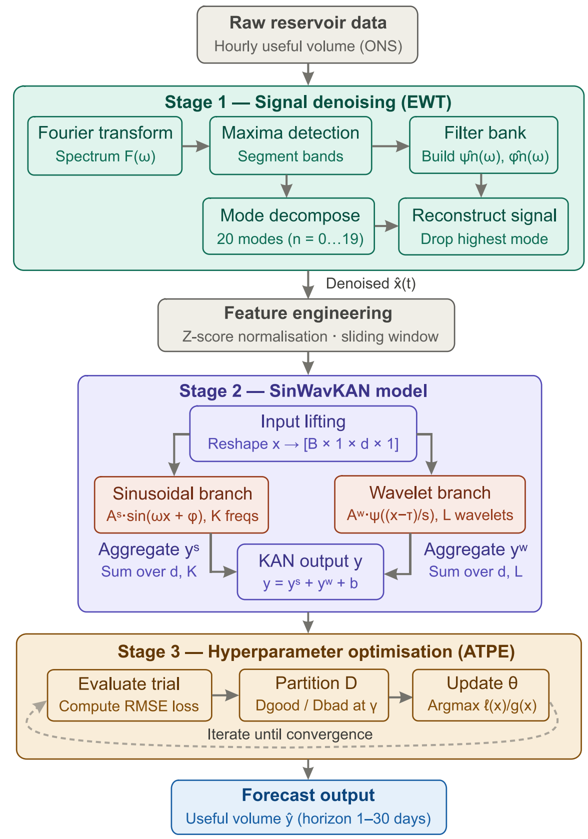

Fig. 2. Proposed method pipeline.

The ONS is responsible for the reservoirs' running during the hydropower plant's regular operations. Under alert conditions, the operation supervisor is in charge of the reservoir's control, and in emergency situations, the plant management is in charge of the reservoir's operation and control. When there are critical flood inflows or when the safety of the plant's facilities or the surrounding area is at risk, the management must establish operational instructions for carrying out the operations [25].

The difference in water flow from the dam, considering the accumulated flow (F_{Acc}), which is determined by the difference between the affluent flow (F_{Aff}) and the defluent flow (F_{Def}), provides the basis for controlling the water level of the hydroelectric plant, calculated according to:

$$F_{\mathrm{Acc}}(t) = F_{\mathrm{Aff}}(t) - F_{\mathrm{Def}}(t). \tag{5}$$

The dam's level rises when $F_{\mathrm{Aff}} > F_{\mathrm{Def}}$ due to the difference in water flow through the plant. Downstream flow (F_{Dow}), which is the turbine flow (F_{Tur}) added to spilled flow (F_{Spi}), is produced by the F_{Def}, as follows:

$$F_{\mathrm{Dow}}(t) = F_{\mathrm{Tur}}(t) + F_{\mathrm{Spi}}(t). \tag{6}$$

Given that the upstream flow (F_{Ups}) is connected to the affluent flow, the water drop (F_{Wat}) is a metric to measure when the plant has the capacity to turbine the water and thus generate energy, given by:

$$F_{\mathrm{Wat}}(t) = F_{\mathrm{Ups}}(t) - F_{\mathrm{Dow}}(t). \tag{7}$$

The flows cause the dam's level to vary with respect to its capacity, which is also referred to as the plant's useful volume. Utilizing Eqs. (5), (6), and (7), the variation in the dam's useful volume is determined [38]. For all of the analyses in this paper, the useful volume of each power plant is the data used in this paper.

4. Methodology

In this section, the methodology of the proposed Opt-EWT-SinWavKAN is explained. Given the problem definition concerning how the volume is considered in the Brazilian electrical power system, as previously explained in Section 3. The proposed hybrid model is presented here (see complete pipeline in Fig. 2), where the EWT (Section 4.1), SinWavKAN (Section 4.2), and ATPE (Section 4.3) are explained.

In the proposed method, preprocessing is initially carried out to decompose the signal. After this initial stage, the proposed Harmonic-Multiscale Kolmogorov–Arnold Network, in short SinWavKAN model, is applied to predict the useful volume of hydroelectric plants. More details on the techniques used to build this structure are presented here.

4.1. Data preprocessing and feature engineering (EWT)

The empirical Wavelets process begins by segmenting the signal's Fourier spectrum into contiguous frequency bands. This segmentation is achieved through the detection of local maxima in the spectrum, which indicates the presence of dominant frequency bands [42]. Once the spectrum is partitioned, empirical Wavelets are constructed as bandpass filters for each segment. These filters are designed to capture the features of the signal within their respective frequency bands, enabling a representation of the signal's intrinsic modes [43].

Empirical scaling $\hat{\phi}_n(\omega)$ and the empirical Wavelets $\hat{\psi}_n(\omega)$ functions are given by:

$$\hat{\phi}_n(\omega) = \begin{cases} 1, & \text{if } |\omega| \le \omega_n - \tau_n, \\ \cos\left(\frac{\pi}{2}\beta\left(\frac{|\omega|-\omega_n+\tau_n}{2\tau_n}\right)\right), & \text{if } \omega_n - \tau_n \le |\omega| \le \omega_n + \tau_n, \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

$$\hat{\psi}_n(\omega) = \begin{cases} 1, & \text{if } \omega_n + \tau_n \le |\omega| \le \omega_{n+1} - \tau_{n+1}, \\ \cos\left(\frac{\pi}{2}\beta\left(\frac{|\omega|-\omega_{n+1}+\tau_{n+1}}{2\tau_{n+1}}\right)\right), & \text{if } \omega_{n+1} - \tau_{n+1} \le |\omega| \le \omega_{n+1} + \tau_{n+1}, \\ \sin\left(\frac{\pi}{2}\beta\left(\frac{|\omega|-\omega_n+\tau_n}{2\tau_n}\right)\right), & \text{if } \omega_n - \tau_n \le |\omega| \le \omega_n + \tau_n, \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

where the $\beta(x)$ is an arbitrary function, so that:

$$\beta(x) = \begin{cases} 0, & \text{if } x \le 0, \\ 1, & \text{if } x \ge 1, \end{cases} \quad \text{and } \beta(x) + \beta(1-x) = 1, \forall x \in [0,1]. \tag{10}$$

Choosing τ_n proportional to ω_n : $\tau_n = \gamma\omega_n$ where $0 < \gamma < 1$, then $\forall > 0$, and Eqs. (8) and (9) can be simplified to:

$$\hat{\phi}_n(\omega) = \begin{cases} 1, & \text{if } |\omega| \le (1-\gamma)\omega_n, \\ \cos\left(\frac{\pi}{2}\beta\left(\frac{|\omega|-(1-\gamma)\omega_n}{2\gamma\omega_n}\right)\right), & \text{if } (1-\gamma)\omega_n \le |\omega| \le (1+\gamma)\omega_n, \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

$$\hat{\psi}_n(\omega) = \begin{cases} 1, & \text{if } (1+\gamma)\omega_n \le |\omega| \le (1-\gamma)\omega_{n+1}, \\ \cos\left(\frac{\pi}{2}\beta\left(\frac{|\omega|-(1-\gamma)\omega_{n+1}}{2\gamma\omega_{n+1}}\right)\right), & \text{if } (1-\gamma)\omega_{n+1} \le |\omega| \le (1+\gamma)\omega_{n+1}, \\ \sin\left(\frac{\pi}{2}\beta\left(\frac{|\omega|-(1-\gamma)\omega_n}{2\gamma\omega_n}\right)\right), & \text{if } (1-\gamma)\omega_n \le |\omega| \le (1+\gamma)\omega_n, \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

Based on $\hat{\psi}_n(\omega)$ and $\hat{\phi}_n(\omega)$, the Empirical Wavelet Transform (EWT) ($W_f^E(n,t)$) can be defined [44]. The coefficients are determined by the internal products of the empirical Wavelets, according to:

$$W_f^E(n,t) = \langle f, \psi_n \rangle = \int f(\tau)\overline{\psi_n(\tau - t)}\, d\tau, \tag{13}$$

$$= \left(\hat{f}(\omega)\overline{\hat{\psi}_n(\omega)}\right)^{\vee}. \tag{14}$$

and the scaling function's inner product with the approximation coefficients is given by:

$$W_f^E(0,t) = \langle f, \phi_1 \rangle = \int f(\tau)\overline{\phi_1(\tau - t)}\, d\tau, \tag{15}$$

$$= \left(\hat{f}(\omega)\overline{\hat{\phi}_1(\omega)}\right)^{\vee}, \tag{16}$$

where $\hat{\psi}_n(\omega)$ and $\hat{\phi}_n(\omega)$ are given by Eqs. (11) and (12). The reconstruction is obtained by:

$$f(t) = W_f^E(0,t) * \phi_1(t) + \sum_{n=1}^{N} W_f^E(n,t) * \psi_n(t) \tag{17}$$

$$= \left(\widehat{W_f^E}(0,\omega)\hat{\phi}_1(\omega) + \sum_{n=1}^{N} \widehat{W_f^E}(n,\omega)\hat{\psi}_n(\omega)\right)^{\vee}. \tag{18}$$

The empirical mode f_k is given by:

$$f_0(t) = W_f^E(0,t) * \phi_1(t), \tag{19}$$

$$f_k(t) = W_f^E(k,t) * \psi_k(t). \tag{20}$$

Algorithm 1 details the implementation of the EWT process applied to our reservoir data. The algorithm employs scale-space representation for robust maxima detection in the Fourier domain, followed by adaptive filter bank construction with optimized transition bands (γ parameter). For each of the parameters (useful volume, inflow, and natural energy inflow), the EWT was applied using a significance threshold θ to discriminate between informative modes and noise components. All numerical features are standardized using z-score normalization: $x_{norm} = \frac{x-\bar{x}}{\sigma}$, where $\bar{x}$ (mean of x values) and σ (standard deviation) are calculated per feature across the training dataset [45].

In this paper, the EWT is applied using the `EWT1D` function available in the `ewtpy` library. The EWT decomposes the original time series of natural flow into N adaptive frequency subbands, which is the hyperparameter used to define the smoothing level. For the denoising procedure, the highest frequency component is excluded from the reconstruction process, and the denoised signal is then obtained by summing the remaining $N - 1$ components.

The EWT decomposition was applied to the complete dataset prior to model training because it is used as a deterministic signal decomposition step rather than a learned transformation. The objective of this preprocessing stage is to separate the signal into frequency bands and remove high-frequency noise components, which depend on the global spectral characteristics of the reservoir dynamics rather than on the prediction targets.

Algorithm 1 EWT for reservoir data filtering.

Input: Time series data $x(t)$ for each reservoir parameter (useful volume, inflow, natural energy inflow)
Output: Filtered time series data $\hat{x}(t)$

/* Phase 1: Fourier spectrum segmentation */
$\hat{x}(\omega) \leftarrow \mathcal{F}[x(t)]$; /* Compute Fourier transform */
$|\hat{x}(\omega)| \leftarrow$ magnitude of $\hat{x}(\omega)$ $\;\Omega \leftarrow \{|\hat{x}(\omega)| : \omega \in [0,\pi]\}$; /* Normalized frequency range */
$\{\omega_i\}_{i=1}^{N} \leftarrow$ DetectLocalMaxima(Ω) ; /* Scale-space approach */

/* Phase 2: Boundary detection */
for $i \leftarrow 1$ **to** $N-1$ **do**
  $\omega_i^{mid} \leftarrow \frac{\omega_i + \omega_{i+1}}{2}$; /* Midpoint between adjacent maxima */
  $\Omega_i \leftarrow [\omega_{i-1}^{mid}, \omega_i^{mid}]$; /* Define frequency support */
end

/* Phase 3: Build empirical Wavelet filter bank */
for $i \leftarrow 1$ **to** N **do**
  $\psi_i(\omega) \leftarrow$ BuildTransitionFunction(ω, Ω_i, γ) ; /* γ is transition width */
end

/* Phase 4: Signal decomposition and filtering */
for $i \leftarrow 1$ **to** N **do**
  $W_f^{\epsilon}(i,t) \leftarrow \mathcal{F}^{-1}[\hat{x}(\omega) \cdot \psi_i(\omega)]$; /* Extract empirical mode */
end

/* Phase 5: Reconstruction with selected modes */
$S \leftarrow$ SelectSignificantModes($\{W_f^{\epsilon}(i,t)\}_{i=1}^{N}$, θ) ; /* θ is significance threshold */
$\hat{x}(t) \leftarrow \sum_{i \in S} W_f^{\epsilon}(i,t)$; /* Reconstructed filtered signal */
return $\hat{x}(t)$

Since EWT does not use label information and does not involve parameter learning related to the forecasting task, it was treated as a signal conditioning procedure similar to filtering operations commonly applied in hydrological time series analysis. The decomposition is fixed before model training and does not adapt based on model performance, which reduces the risk of indirect information leakage. Nevertheless, this design choice aims to ensure stable frequency band identification across the entire observation period and to avoid inconsistencies that could arise from independently decomposing shorter temporal segments.

4.2. Harmonic-multiscale Kolmogorov-arnold network (SinWavKAN)

We propose the Harmonic-Multiscale Kolmogorov–Arnold Network, a KAN architecture in which each univariate edge function is parameterized by a combination of harmonic spectral components by Sinusoidal expansion and multiscale Wavelets (Mexican hat Wavelet), called in short as SinWavKAN. The architecture of the proposed SinWavKAN is presented in Fig. 3. The harmonic component provides a flexible spectral representation that subsumes Sinusoidal expansions while remaining agnostic to a specific trigonometric parameterization, enabling smooth modeling of global oscillatory structure.

The multiscale component introduces localized basis functions with learnable scales and translations, allowing the network to capture non-stationary and spatially localized variations across multiple resolutions, a principle well established in signal processing theory. By embedding these representations within the Kolmogorov–Arnold framework, the model preserves the additive structure of univariate edge functions, ensuring interpretability and theoretical grounding while significantly enhancing expressive capacity.

According to the Kolmogorov–Arnold representation theorem, any continuous multivariate function ($f : [0,1]^d \to \mathbb{R}$) can be expressed as a finite superposition of univariate nonlinear functions and linear combinations. Modern KANs operationalize this principle by replacing fixed nonlinearities with learnable univariate functions acting on individual

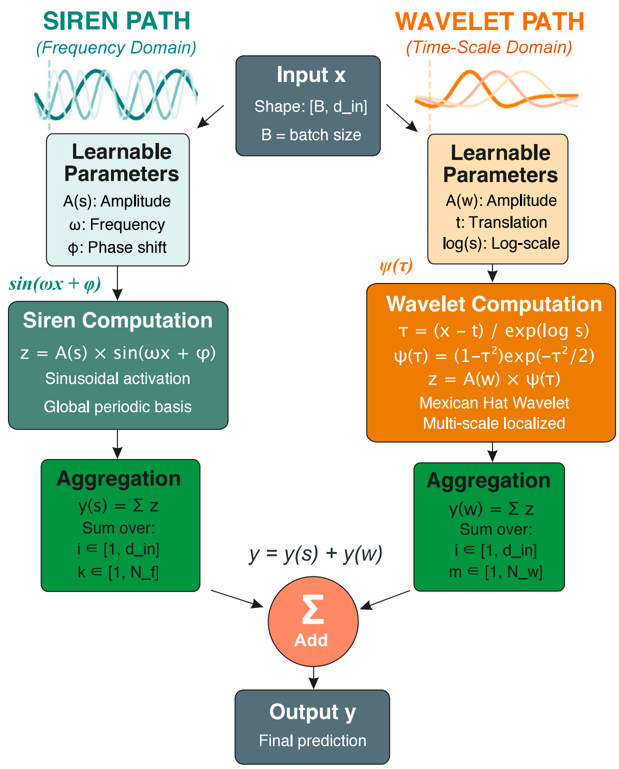

Fig. 3. SinWavKAN architecture: A Harmonic-Multiscale Kolmogorov–Arnold Network.

input coordinates [46]. In this framework, each output component of a KAN is written as

$$f_j(\mathbf{x}) = \sum_{i=1}^{d} \phi_{j,i}(x_i) + b_j, \tag{21}$$

where $\phi_{j,i} : \mathbb{R} \to \mathbb{R}$ are learnable univariate functions and b_j is a bias term.

The proposed SinWavKAN model instantiates each $\phi_{j,i}$ as an explicit parametric expansion combining Sinusoidal and Wavelet bases, yielding an analytically tractable and interpretable realization of KANs. For each output dimension j and input dimension i, the associated univariate function $\phi_{j,i}$ is defined as

$$\phi_{j,i}(x) = \phi^{\text{sin}}_{j,i}(x) + \phi^{\text{wav}}_{j,i}(x), \tag{22}$$

where $\phi^{\text{sin}}_{j,i}$ and $\phi^{\text{wav}}_{j,i}$ denote Sinusoidal and Wavelet components, respectively. This additive decomposition aligns with the KAN philosophy of explicitly modeling and visualizing each univariate interaction.

4.2.1. Sinusoidal expansion within KANs

The Sinusoidal component is defined as

$$\phi^{\text{sin}}_{j,i}(x) = \sum_{k=1}^{K} A^{\text{sin}}_{j,i,k} \sin\left(\omega_{j,i,k} x + \phi_{j,i,k}\right), \tag{23}$$

where $K = n_{\text{freq}}$ is the number of Sinusoidal atoms.

From a KAN perspective, this expansion can be interpreted as a learned Fourier-like basis for each univariate function. Unlike classical KAN implementations based on splines or piecewise polynomials, the Sinusoidal parameterization provides a smooth, globally supported basis with explicitly learnable frequencies and phases. The derivative of $\phi^{\text{sin}}_{j,i}$ with respect to the input is given by

$$\frac{d}{dx}\phi^{\text{sin}}_{j,i}(x) = \sum_{k=1}^{K} A^{\text{sin}}_{j,i,k}\omega_{j,i,k} \cos\left(\omega_{j,i,k} x + \phi_{j,i,k}\right), \tag{24}$$

which remains bounded and non-vanishing for a wide range of frequencies, supporting stable gradient propagation during training.

4.2.2. Wavelet expansion as local KAN basis

To complement the global support of Sinusoidal functions, the Wavelet component introduces localized basis functions. Using the Mexican hat Wavelet ψ, the Wavelet component is defined as

$$\phi^{\text{wav}}_{j,i}(x) = \sum_{l=1}^{L} A^{\text{wav}}_{j,i,l}\, \psi\left(\frac{x - \tau_{j,i,l}}{s_{j,i,l}}\right), \tag{25}$$

with $L = n_{\text{Wavelets}}$.

The Mexican Hat Wavelet

$$\psi(t) = (1 - t^2) \exp\left(-\tfrac{1}{2}t^2\right) \tag{26}$$

satisfies $\int_{\mathbb{R}} \psi(t)\,dt = 0$, which ensures sensitivity to local variations rather than constant offsets. The scale parameters are constrained to be positive via $s_{j,i,l} = \exp(\sigma_{j,i,l})$, which ensures differentiability and numerical stability.

4.2.3. Unified KAN mapping

Substituting the univariate expansions into the KAN formulation yields

$$f_j(\mathbf{x}) = \sum_{i=1}^{d} \left[\sum_{k=1}^{K} A^{\text{sin}}_{j,i,k} \sin(\omega_{j,i,k} x_i + \phi_{j,i,k}) + \sum_{l=1}^{L} A^{\text{wav}}_{j,i,l}\, \psi\left(\frac{x_i - \tau_{j,i,l}}{s_{j,i,l}}\right) \right] + b_j. \tag{27}$$

This expression explicitly satisfies the Kolmogorov–Arnold structural constraint, as all nonlinearities are univariate and all multivariate interactions arise solely through summation. The Sinusoidal basis provides dense coverage of smooth function spaces, while the Wavelet basis enables sparse representation of localized irregularities. Their joint use allows each $\phi_{j,i}$ to adapt its smoothness locally, which is not possible with spline-based KANs using a fixed knot structure. The pseudocode of the SinWaveKAN is presented in Algorithm 2.

Under mild regularity assumptions, the combined basis forms a universal approximator for continuous univariate functions on compact domains. Consequently, the full SinWavKAN model inherits universal approximation properties for multivariate functions through the Kolmogorov–Arnold construction.

Each learned function $\phi_{j,i}$ can be directly visualized and analyzed, preserving one of the central advantages of KANs. The separation between Sinusoidal and Wavelet components further allows decomposition of learned relationships into global oscillatory behavior and localized features. SinWavKAN realizes a Kolmogorov–Arnold Network in which univariate edge functions are parameterized by adaptive spectral and multiscale bases. This design preserves the theoretical guarantees of KANs while significantly enhancing their expressive capacity for nonstationary and multiscale phenomena.

4.3. Adaptive tree-structured parzen estimator

Adaptive Tree-structured Parzen Estimator, abbreviated as ATPE, is a sequential model-based optimization method designed to solve black box optimization problems of the form

$$\mathbf{x}^\star = \arg\min_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x}), \tag{28}$$

where $\mathcal{X} \subset \mathbb{R}^d$ denotes the search space and $f(\mathbf{x})$ is an expensive to evaluate objective function.

ATPE builds upon the tree-structured Parzen Estimator framework by introducing adaptive mechanisms that dynamically adjust the sampling strategy according to the observed performance of previous evaluations. The algorithm models the conditional probability distribution of the hyperparameters given the objective value, expressed as $p(\mathbf{x} \mid y)$, where $y = f(\mathbf{x})$. The observed set of evaluations $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$ is

Algorithm 2 SinWaveKAN forward propagation

Input: Input signal $x \in \mathbb{R}^{B\times d}$, number of sinusoidal components N_f, number of wavelet components N_w
Output: Network output $y \in \mathbb{R}^{B\times o}$

```
/* Phase 1: Input lifting */
```
$\tilde{x} \leftarrow \mathrm{Reshape}(x, B\times 1\times d\times 1)$
```
/* Phase 2: Sinusoidal (SIREN) basis expansion */
```
for $i \leftarrow 1$ **to** o **do**
  for $j \leftarrow 1$ **to** d **do**
    for $k \leftarrow 1$ **to** N_f **do**
      $\phi_{ijk}(\tilde{x}_j) \leftarrow a^{s}_{ijk}\sin(\omega_{ijk}\tilde{x}_j + \varphi_{ijk})$
    end
  end
end
$y^{s}_i \leftarrow \sum_{j=1}^{d}\sum_{k=1}^{N_f}\phi_{ijk}(\tilde{x}_j)$
```
/* Phase 3: Wavelet (Mexican hat) basis expansion */
```
$s_{ijk} \leftarrow \exp(\ell_{ijk})$; /* Positive scale parameter */
for $i \leftarrow 1$ **to** o **do**
  for $j \leftarrow 1$ **to** d **do**
    for $k \leftarrow 1$ **to** N_w **do**
      $t_{ijk} \leftarrow \dfrac{\tilde{x}_j - \tau_{ijk}}{s_{ijk}}$ $\psi_{ijk}(\tilde{x}_j) \leftarrow a^{w}_{ijk}\left(1 - t^2_{ijk}\right)\exp\left(-\frac{1}{2}t^2_{ijk}\right)$
    end
  end
end
$y^{w}_i \leftarrow \sum_{j=1}^{d}\sum_{k=1}^{N_w}\psi_{ijk}(\tilde{x}_j)$
```
/* Phase 4: KAN aggregation */
```
$y_i \leftarrow y^{s}_i + y^{w}_i$ **if** *bias is enabled* **then**
  $y_i \leftarrow y_i + b_i$
end
return y

partitioned using a performance threshold y^*, typically defined as a quantile $\gamma \in (0,1)$ of the observed objective values. This induces two disjoint subsets

$$\mathcal{D}_{\text{good}} = \{\mathbf{x}_i \mid y_i \le y^*\}, \tag{29}$$

$$\mathcal{D}_{\text{bad}} = \{\mathbf{x}_i \mid y_i > y^*\}. \tag{30}$$

Using these subsets, ATPE constructs two nonparametric density estimators

$$\ell(\mathbf{x}) = p(\mathbf{x} \mid y \le y^*), \tag{31}$$

$$g(\mathbf{x}) = p(\mathbf{x} \mid y > y^*), \tag{32}$$

which are commonly modeled using Parzen window estimators with adaptive kernel bandwidths.

The optimization objective is reformulated by maximizing the expected improvement [47], which under the TPE formulation is equivalent to maximizing the ratio

$$\mathbf{x}^\star = \arg\max_{\mathbf{x}} \frac{\ell(\mathbf{x})}{g(\mathbf{x})}. \tag{33}$$

ATPE extends this formulation by dynamically adapting several internal parameters of the estimator during the optimization process. These adaptations include the quantile parameter γ, the kernel bandwidths of the Parzen estimators, and the sampling weights associated with individual dimensions of $\mathbf{x}$. The adaptive behavior is driven by empirical statistics extracted from $\mathcal{D}$, such as the variance of objective values and the relative importance of each hyperparameter [48].

Let θ denote the set of internal modeling parameters of the estimator. ATPE updates θ at each iteration t according to

$$\theta^{(t+1)} = \mathcal{A}(\theta^{(t)}, \mathcal{D}^{(t)}), \tag{34}$$

where $\mathcal{A}(\cdot)$ represents an adaptive update rule learned from previous optimization trajectories. By continuously refining the probabilistic model and its hyperparameters, ATPE achieves a more efficient exploration–exploitation trade-off compared to static TPE, leading to faster convergence and improved robustness in high-dimensional and heterogeneous search spaces [49].

5. Results

In this section, we present the results of applying our proposed model. First, an ablation study is presented to prove that the choice of each technique is worth it, and well combined to achieve the best hybrid architecture considering the raw signal. After that, we present the results of signal denoising and hypertuning. Based on an optimized architecture, we present an evaluation of the horizon variation and window size to feed our prediction model, followed by a benchmarking and generalization analysis, and a statistical assessment.

5.1. Experimental setup

The experimental evaluation was conducted on a computational environment (`Google Colab`) designed to ensure reproducibility and adequate performance for training and validation of the proposed models. The system was equipped with 12.7 GB of system random access memory and 112.6 GB of available disk storage, which were sufficient to handle the dataset, intermediate artifacts, and model checkpoints generated during the experiments. All models were trained and evaluated using an NVIDIA T4 graphics processing unit.

This hardware configuration was kept fixed across all experiments to guarantee fair comparison among models in terms of predictive accuracy, computational cost, and model complexity. The number of parameters corresponds to the total count of trainable weights in each model, while the training time and validation time, expressed in seconds, quantify the computational efficiency of the learning and evaluation processes on the adopted hardware.

The model is constructed using a sliding window strategy with a multiple-step-ahead forecasting horizon, using an 80% training and 20% testing split. The architectural configuration is defined by the number of sinusoidal components (`n_freq`) and wavelet components (`n_wavelets`), both treated as tunable structural hyperparameters, together with an optional bias term.

The training process is conducted using a supervised learning procedure implemented in `PyTorch`, where model parameters are optimized through gradient-based learning. The hyperparameter optimization is performed using the ATPE, considering a search space defined by `n_freq` $\in [4, 64]$, `n_wavelets` $\in [4, 64]$ with increments of 4, and a binary bias inclusion parameter. A total of 50 optimization trials are conducted, where each candidate configuration is evaluated based on the Root Mean Square Error (RMSE) computed on the test set. The optimization objective is defined to minimize this error, ensuring a fair comparison between architectural configurations.

5.1.1. Evaluation metrics

Model performance was assessed using standard regression metrics, namely the RMSE, the Mean Absolute Error (MAE), and the Symmetric Mean Absolute Percentage Error (SMAPE), together with the number of trainable parameters and the computational time required for training and validation.

RMSE emphasizes large prediction errors due to the quadratic penalty, making it suitable for detecting significant deviations. MAE measures the average absolute error and provides an interpretable assessment of overall prediction performance with lower sensitivity to outliers. SMAPE evaluates the relative prediction error in percentage form while maintaining symmetry between overestimation and underestimation, enabling fair comparisons across different signal magnitudes [49].

Table 2
Ablation study for Fourier, radial basis, Chebyshev, Sinusoidal representation, Wavelet, and ReLU activation functions.

F	R	C	S	W	RL	RMSE	MAE	SMAPE (%)	Params	Training time (s)	Inference time (s)
✓	×	×	×	×	×	3.53×10^{-1}	2.39×10^{-1}	7.82×10^{1}	1.61×10^{2}	9.61×10^{-1}	1.82×10^{-3}
×	✓	×	×	×	×	1.65×10^{-1}	1.18×10^{-1}	4.44×10^{1}	2.41×10^{2}	1.63	1.57×10^{-3}
×	×	✓	×	×	×	3.94×10^{-1}	2.76×10^{-1}	9.33×10^{1}	$\mathbf{4.10}\times10^{1}$	$\mathbf{4.70}\times10^{-1}$	$\mathbf{1.12}\times10^{-3}$
×	×	×	✓	×	×	1.52×10^{-1}	9.81×10^{-2}	3.74×10^{1}	2.41×10^{2}	9.62×10^{-1}	1.30×10^{-3}
×	×	×	×	✓	×	1.69×10^{-1}	1.25×10^{-1}	4.46×10^{1}	2.41×10^{2}	2.67	2.36×10^{-3}
×	×	×	×	×	✓	1.94×10^{-1}	1.46×10^{-1}	5.47×10^{1}	1.61×10^{2}	7.44×10^{-1}	9.47×10^{-4}
✓	✓	×	×	×	×	3.42×10^{-1}	2.20×10^{-1}	6.61×10^{1}	4.01×10^{2}	2.45	2.57×10^{-3}
✓	×	✓	×	×	×	3.29×10^{-1}	2.19×10^{-1}	6.50×10^{1}	2.01×10^{2}	1.35	2.62×10^{-3}
✓	×	×	✓	×	×	3.91×10^{-1}	2.53×10^{-1}	7.54×10^{1}	4.01×10^{2}	2.16	3.19×10^{-3}
✓	×	×	×	✓	×	3.31×10^{-1}	2.14×10^{-1}	6.99×10^{1}	4.01×10^{2}	3.85	3.56×10^{-3}
✓	×	×	×	×	✓	3.03×10^{-1}	1.96×10^{-1}	6.42×10^{1}	3.21×10^{2}	1.59	2.34×10^{-3}
×	✓	✓	×	×	×	2.19×10^{-1}	1.41×10^{-1}	5.37×10^{1}	2.81×10^{2}	1.87	2.63×10^{-3}
×	✓	×	✓	×	×	1.45×10^{-1}	9.59×10^{-2}	4.00×10^{1}	4.81×10^{2}	2.53	2.59×10^{-3}
×	✓	×	×	✓	×	2.07×10^{-1}	1.47×10^{-1}	5.19×10^{1}	4.81×10^{2}	4.49	3.64×10^{-3}
×	✓	×	×	×	✓	1.70×10^{-1}	1.15×10^{-1}	4.51×10^{1}	4.01×10^{2}	2.16	2.39×10^{-3}
×	×	✓	✓	×	×	1.82×10^{-1}	1.11×10^{-1}	4.40×10^{1}	2.81×10^{2}	1.47	2.39×10^{-3}
×	×	✓	×	✓	×	1.88×10^{-1}	1.20×10^{-1}	4.46×10^{1}	2.81×10^{2}	3.31	3.18×10^{-3}
×	×	✓	×	×	✓	1.92×10^{-1}	1.25×10^{-1}	4.79×10^{1}	2.01×10^{2}	1.21	2.04×10^{-3}
×	×	×	✓	✓	×	$\mathbf{1.25}\times10^{-1}$	$\mathbf{8.42}\times10^{-2}$	$\mathbf{3.39}\times10^{1}$	4.81×10^{2}	4.05	3.98×10^{-3}
×	×	×	✓	×	✓	2.02×10^{-1}	1.38×10^{-1}	5.05×10^{1}	4.01×10^{2}	1.82	2.14×10^{-3}
×	×	×	×	✓	✓	2.27×10^{-1}	1.67×10^{-1}	5.69×10^{1}	4.01×10^{2}	3.34	3.04×10^{-3}
✓	✓	✓	×	×	×	3.71×10^{-1}	2.37×10^{-1}	7.05×10^{1}	4.41×10^{2}	2.83	3.90×10^{-3}
✓	✓	×	✓	×	×	2.88×10^{-1}	1.82×10^{-1}	5.66×10^{1}	6.41×10^{2}	3.48	4.70×10^{-3}
✓	✓	×	×	✓	×	1.80×10^{-1}	1.21×10^{-1}	5.13×10^{1}	6.41×10^{2}	5.23	5.07×10^{-3}
✓	✓	×	×	×	✓	3.30×10^{-1}	2.30×10^{-1}	7.30×10^{1}	5.61×10^{2}	3.02	3.50×10^{-3}
✓	×	✓	✓	×	×	3.12×10^{-1}	2.03×10^{-1}	6.32×10^{1}	4.41×10^{2}	2.38	3.37×10^{-3}
✓	×	✓	×	✓	×	2.82×10^{-1}	1.79×10^{-1}	6.38×10^{1}	4.41×10^{2}	4.54	4.62×10^{-3}
✓	×	✓	×	×	✓	3.40×10^{-1}	2.28×10^{-1}	7.11×10^{1}	3.61×10^{2}	2.08	3.17×10^{-3}
✓	×	×	✓	✓	×	1.89×10^{-1}	1.30×10^{-1}	4.75×10^{1}	6.41×10^{2}	4.58	4.71×10^{-3}
✓	×	×	✓	×	✓	2.25×10^{-1}	1.35×10^{-1}	5.10×10^{1}	5.61×10^{2}	2.68	5.13×10^{-3}
✓	×	×	×	✓	✓	2.63×10^{-1}	1.68×10^{-1}	6.28×10^{1}	5.61×10^{2}	4.81	4.46×10^{-3}

Given a set of ground truth values $\{y_i\}_{i=1}^{N}$ and corresponding predictions $\{\hat{y}_i\}_{i=1}^{N}$, these metrics are defined as

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}, \tag{35}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right|, \tag{36}$$

$$\text{SMAPE} = \frac{100}{N}\sum_{i=1}^{N}\frac{\left|y_i - \hat{y}_i\right|}{\left(\left|y_i\right| + \left|\hat{y}_i\right|\right)/2}. \tag{37}$$

5.1.2. Compared models

To evaluate the effectiveness of the proposed model, comparisons are conducted against several variants of KAN based architectures, including the Standard KAN (StandardKAN) [50], the Fast KAN (fastKAN) [51], the Efficient KAN (efficientKAN) [52], the Rectified Linear Unit based KAN (ReLUKAN), the Sinusoidal Representation KAN (SirenKAN), the Chebyshev Polynomial based KAN (ChebyKAN) [53], the Wavelet based KAN (WavKAN), the Radial Basis Function KAN (RBFKAN) [54], and the Fourier series based KAN (FourierKAN) [55]. These models were selected because they represent distinct functional basis expansions and architectural adaptations of the Kolmogorov–Arnold framework, providing a comprehensive and representative cooperative analysis for assessing the performance and computational efficiency of the proposed approach.

To conduct benchmarking, we compare the proposed Opt-EWT-SinWavKAN with state-of-the-art models from different architecture families. These include Multilayer Perceptron-based models, such as Time series Dense Encoder (TiDE) [56], Neural Hierarchical Interpolation for Time Series Forecasting (NHITS) [57], Neural Basis Expansion Analysis for Time Series Forecasting (N-BEATS) [58], N-BEATS extended with exogenous variables (N-BEATSx) [59], Time-Series Mixer (TSMixer), and TSMixerx.

We also consider RNN-based models, including Dilated RNN (DilatedRNN) [60] and Vanilla RNN (VanillaRNN), Long Short Term Memory networks (LSTM) [61], Gated Recurrent Units (GRU) [62], and Deep Autoregressive models (DeepAR) [63]. In addition, Transformer-based models are evaluated, such as the Temporal Fusion Transformer (TFT) [64], Informer [65], iTransformer [66], Patch Time Series Transformer (PatchTST) [67], and Frequency Enhanced Decomposed Transformer (FEDformer) [68]. Finally, we include Convolutional Neural Network-based models, namely the Temporal Convolutional Network (TCN) [69], Bidirectional TCN (BiTCN) [70], and TimesNet [71].

5.1.3. Dataset

The FUNDAO hydroelectric plant was considered for this study, serving as the basis for comparative analyses. In addition, the SEGREDO and FOZ DO CHAPECÓ plants were considered for a generalization analysis of the model proposed in this paper. The data from these three hydroelectric power plants is publicly available at: https://dados.ons.org.br/dataset/ear-diario-por-reservatorio (accessed on February 06, 2026). The data used in this study considers hourly measurements of `val_volumeutil`, with a total of 19,656 samples corresponding to a period from 2022-09-03, 23:59:00 to 2024-11-30, 23:59:00.

The FUNDAO Hydroelectric Power Plant (HPP), located on the Jordão River, has an installed capacity ranging from approximately 124 MW. Commissioned around 2006, it supplies electricity to the southwestern region of Paraná, between the municipalities of Pinhão and Foz do Jordão. The Governador Ney Aminthas de Barros Braga Hydroelectric Power Plant, also known as SEGREDO HPP, is a facility operated by Copel on the Iguaçu River, with an installed capacity of 1260 MW, providing energy to nearly three million people. The FOZ DO CHAPECÓ HPP is notable for its 598 m long and 48 m high dam with an installed capacity of 855 MW, which was the first in Brazil to adopt an asphalt core technique.

Table 3
Ablation study (continued).

F	R	C	S	W	RL	RMSE	MAE	SMAPE (%)	Params	Training time (s)	Inference time (s)
×	✓	✓	✓	×	×	2.13×10^{-1}	1.44×10^{-1}	5.23×10^{1}	5.21×10^{2}	2.92	3.74×10^{-3}
×	✓	✓	×	✓	×	3.18×10^{-1}	2.02×10^{-1}	6.57×10^{1}	5.21×10^{2}	4.65	4.79×10^{-3}
×	✓	✓	×	×	✓	2.11×10^{-1}	1.33×10^{-1}	4.74×10^{1}	4.41×10^{2}	3.17	3.20×10^{-3}
×	✓	×	✓	✓	×	2.62×10^{-1}	2.04×10^{-1}	6.15×10^{1}	7.21×10^{2}	5.13	4.76×10^{-3}
×	✓	×	✓	×	✓	1.43×10^{-1}	9.20×10^{-2}	3.57×10^{1}	6.41×10^{2}	3.09	3.35×10^{-3}
×	✓	×	×	✓	✓	1.45×10^{-1}	1.00×10^{-1}	3.98×10^{1}	6.41×10^{2}	5.09	4.27×10^{-3}
×	×	✓	✓	✓	×	2.38×10^{-1}	1.48×10^{-1}	5.53×10^{1}	5.21×10^{2}	4.34	4.64×10^{-3}
×	×	✓	✓	×	✓	1.62×10^{-1}	1.04×10^{-1}	3.91×10^{1}	4.41×10^{2}	2.30	3.05×10^{-3}
×	×	✓	×	✓	✓	2.24×10^{-1}	1.46×10^{-1}	5.30×10^{1}	4.41×10^{2}	4.22	5.65×10^{-3}
×	×	×	✓	✓	✓	1.73×10^{-1}	1.05×10^{-1}	4.43×10^{1}	6.41×10^{2}	4.54	4.56×10^{-3}
✓	✓	✓	✓	×	×	3.30×10^{-1}	2.14×10^{-1}	6.38×10^{1}	6.81×10^{2}	3.72	6.74×10^{-3}
✓	✓	✓	×	✓	×	2.71×10^{-1}	1.74×10^{-1}	5.64×10^{1}	6.81×10^{2}	5.82	6.07×10^{-3}
✓	✓	✓	×	×	✓	3.44×10^{-1}	2.19×10^{-1}	6.44×10^{1}	6.01×10^{2}	3.47	4.36×10^{-3}
✓	✓	×	✓	✓	×	2.65×10^{-1}	1.77×10^{-1}	6.36×10^{1}	8.81×10^{2}	6.07	8.29×10^{-3}
✓	✓	×	✓	×	✓	2.34×10^{-1}	1.51×10^{-1}	4.90×10^{1}	8.01×10^{2}	4.38	4.52×10^{-3}
✓	✓	×	×	✓	✓	3.38×10^{-1}	2.16×10^{-1}	6.49×10^{1}	8.01×10^{2}	5.64	6.41×10^{-3}
✓	×	✓	✓	✓	×	2.65×10^{-1}	1.83×10^{-1}	5.84×10^{1}	6.81×10^{2}	5.59	5.57×10^{-3}
✓	×	✓	✓	×	✓	2.46×10^{-1}	1.65×10^{-1}	5.77×10^{1}	6.01×10^{2}	3.17	4.34×10^{-3}
✓	×	✓	×	✓	✓	2.24×10^{-1}	1.55×10^{-1}	5.55×10^{1}	6.01×10^{2}	4.80	5.17×10^{-3}
✓	×	×	✓	✓	✓	3.01×10^{-1}	1.89×10^{-1}	5.95×10^{1}	8.01×10^{2}	5.67	5.40×10^{-3}
×	✓	✓	✓	✓	×	1.99×10^{-1}	1.25×10^{-1}	4.49×10^{1}	7.61×10^{2}	5.51	5.67×10^{-3}
×	✓	✓	✓	×	✓	2.95×10^{-1}	2.07×10^{-1}	6.63×10^{1}	6.81×10^{2}	3.77	5.93×10^{-3}
×	✓	✓	×	✓	✓	1.90×10^{-1}	1.13×10^{-1}	4.46×10^{1}	6.81×10^{2}	5.59	5.11×10^{-3}
×	✓	×	✓	✓	✓	1.71×10^{-1}	1.13×10^{-1}	3.98×10^{1}	8.81×10^{2}	5.82	6.08×10^{-3}
×	×	✓	✓	✓	✓	2.33×10^{-1}	1.49×10^{-1}	5.91×10^{1}	6.81×10^{2}	5.39	5.95×10^{-3}
✓	✓	✓	✓	✓	×	3.35×10^{-1}	2.14×10^{-1}	7.20×10^{1}	9.21×10^{2}	6.50	9.07×10^{-3}
✓	✓	✓	✓	×	✓	2.23×10^{-1}	1.39×10^{-1}	5.20×10^{1}	8.41×10^{2}	5.02	5.52×10^{-3}
✓	✓	✓	×	✓	✓	2.62×10^{-1}	1.68×10^{-1}	5.54×10^{1}	8.41×10^{2}	6.25	6.65×10^{-3}
✓	✓	×	✓	✓	✓	2.35×10^{-1}	1.52×10^{-1}	5.35×10^{1}	1.04×10^{3}	7.21	6.21×10^{-3}
✓	×	✓	✓	✓	✓	3.50×10^{-1}	2.35×10^{-1}	7.33×10^{1}	8.41×10^{2}	5.90	6.61×10^{-3}
×	✓	✓	✓	✓	✓	2.12×10^{-1}	1.43×10^{-1}	4.96×10^{1}	9.21×10^{2}	6.93	5.95×10^{-3}
✓	✓	✓	✓	✓	✓	3.27×10^{-1}	2.04×10^{-1}	6.21×10^{1}	1.08×10^{3}	7.36	7.50×10^{-3}

5.2. Structure design by ablation study

Tables 2 and 3 report a comprehensive ablation study evaluating different combinations of basis functions and activation mechanisms, namely Fourier (F), radial basis functions (R), Chebyshev polynomials (C), Sinusoidal representation networks (S), Wavelets (W), and the ReLU activation (RL). The results reveal a clear dependency of predictive performance on the choice and combination of functional components, as measured by RMSE, MAE, and SMAPE, as well as on model complexity and computational cost.

Considering single component configurations in Table 2, models based solely on Sinusoidal representation and R exhibit the best predictive accuracy among individual bases, with the Sinusoidal representation configuration achieving RMSE = 1.52×10^{-1} and SMAPE = 3.74×10^{1}, outperforming Fourier, Chebyshev, Wavelet, and ReLU alone. In contrast, Chebyshev-only and Fourier-only models present noticeably higher errors, indicating limited expressive power when used in isolation. In terms of efficiency, Chebyshev and ReLU configurations require fewer parameters and shorter training times, highlighting a tradeoff between accuracy and model compactness.

When combining two components, consistent performance gains are observed in several cases. In particular, the radial basis functions combined with the sinusoidal representation significantly improves accuracy compared to either component alone, while maintaining a moderate increase in parameter count. The most notable result in Table 2 is obtained using a combination of Sinusoidal representation and Wavelets, which achieves the best overall performance with RMSE = 1.25×10^{-1}, MAE = 8.42×10^{-2}, and SMAPE = 3.39×10^{1}. This suggests a strong complementarity between Sinusoidal representations and Wavelet bases, where Sinusoidal representation captures global periodic patterns, and W enhances localized time–frequency features. However, this accuracy gain comes at the cost of increased model size and longer training and validation times.

Table 3 extends the analysis to more complex combinations involving three or more components. Overall, adding further bases generally increases the number of parameters and computational cost, while accuracy improvements are not always guaranteed. Although some multi-component configurations yield competitive RMSE and MAE values, none consistently surpass the Sinusoidal representation and Wavelets combination observed in Table 2. In fact, fully combined models incorporating all components tend to suffer from degraded performance, suggesting overparameterization and optimization difficulty. These results indicate that carefully selected hybrid representations can outperform both simple and overly complex architectures.

In summary, the experimental results demonstrate that Sinusoidal-based models play a central role in achieving lower error, particularly when combined with Wavelet representations. While richer combinations increase expressive capacity, the best tradeoff between accuracy, efficiency, and stability is obtained with a limited number of complementary bases rather than exhaustive combinations. Our architecture, based on the use of Sinusoidal representation and Wavelets (the reason why it is called SinWavKAN), is innovative, as it combines two techniques to achieve an enhanced hybrid KAN method.

5.3. Signal denoising analysis

The purpose of noise suppression using EWT is to enable the model to focus on the variation trend, which has a greater influence, rather than on noise. The EWT decomposes the signals in various modes as presented in Fig. 4A for the first 5 modes, considering a total of 20 modes; when they are combined, the result is the original signal. To remove noise from this signal, the highest frequency is disregarded (see Fig. 4B), and thus the filter is implemented. The challenge in applying this filter is to define the level of decomposition necessary for the highest frequency to be representative in relation to noise reduction and not a fundamental part of the signal.

We decompose the signal into 10 to 60 modes and subtract the highest frequency from the decomposition; the results are shown in Fig. 5. When 10 modes are used, the last one is subtracted, meaning that

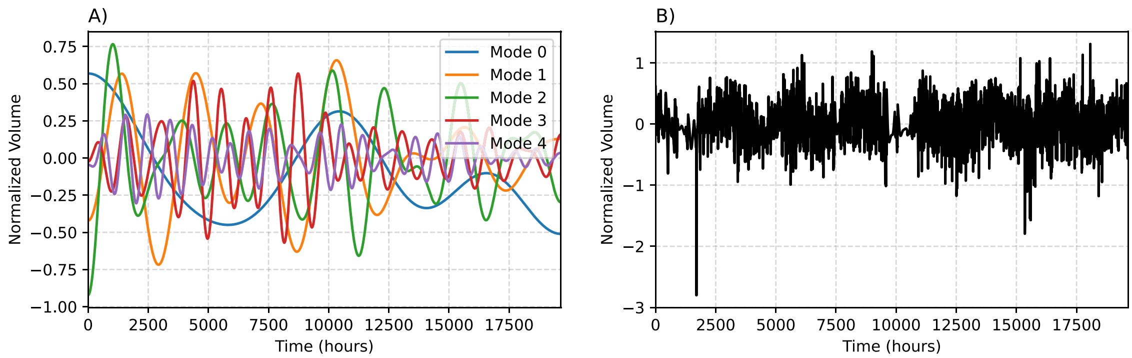

Fig. 4. Signal decomposition: (A) Modes from 0 to 5, (B) Mode 19.

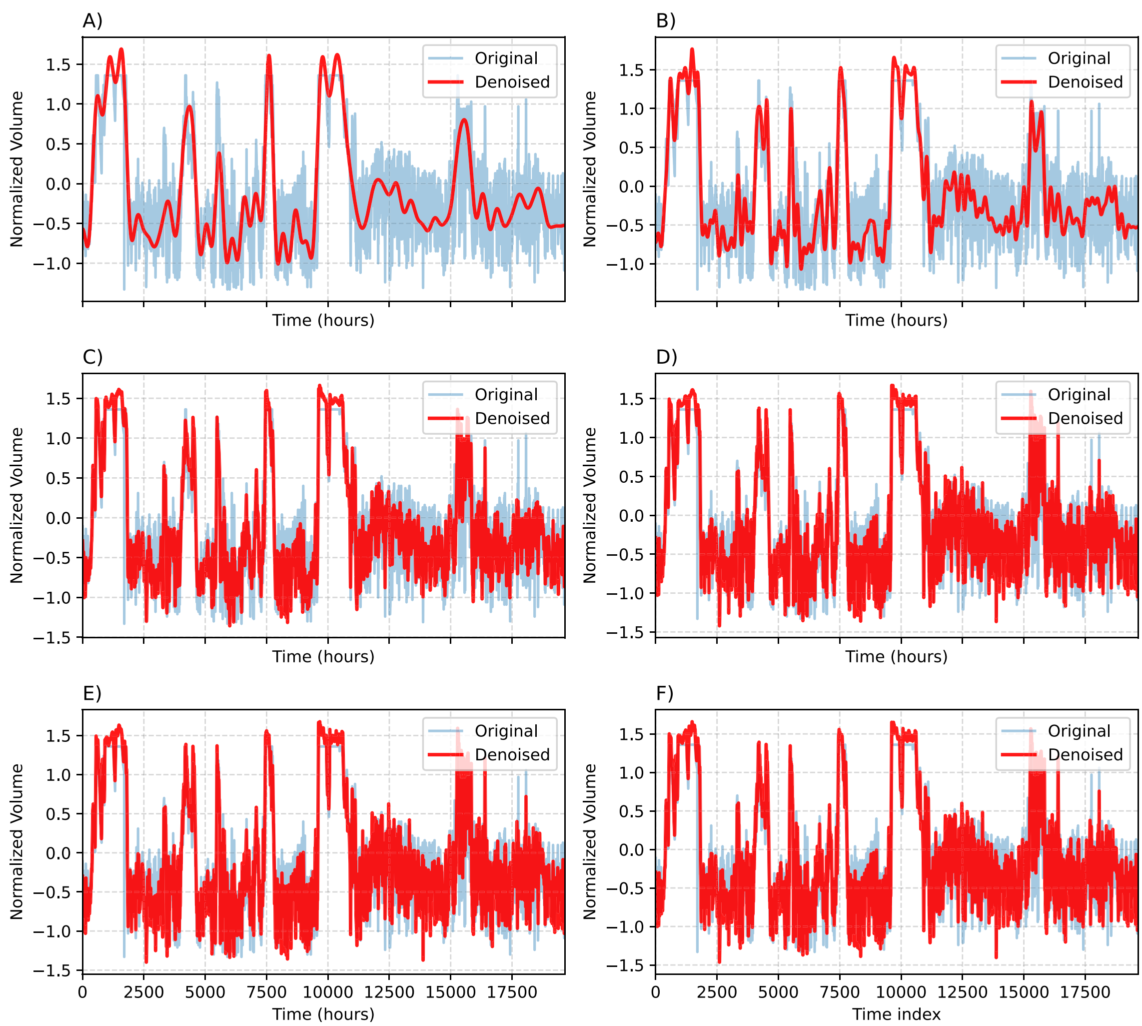

Fig. 5. Denoising based on EWT considering a decomposition using: (A) 10 modes, (B) 20 modes, (C) 30 modes, (D) 40 modes, (E) 50 modes, (F) 60 modes.

the reconstructed signal has 9 modes; in this case, the filter becomes too intense, reducing some fundamental frequencies of the signal, which is not the objective of this decomposition. When the signal is composed of 20 modes, i.e., 19 modes of this decomposition are used to reconstruct the original signal, promising results are achieved, with the signal following the variation trend with acceptable fluctuation, thus reducing the necessary noise. When 30 or more modes are used for decomposition, removing the mode with the highest frequency is not sufficient to attenuate the signal, so the signal remains very close to the original signal. For this reason, 20 modes were used for signal decomposition during filter application.

Proper filter tuning is essential for the model to focus on the signal appropriately. The higher the noise level, the more difficult it will be to predict the variation using the proposed model. Thus, filter tuning

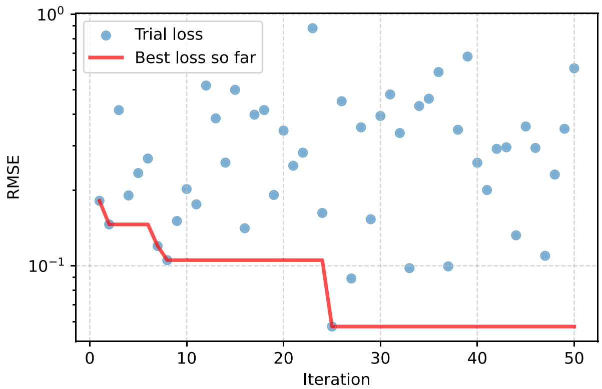

Fig. 6. RMSE per trial during hypertuning.

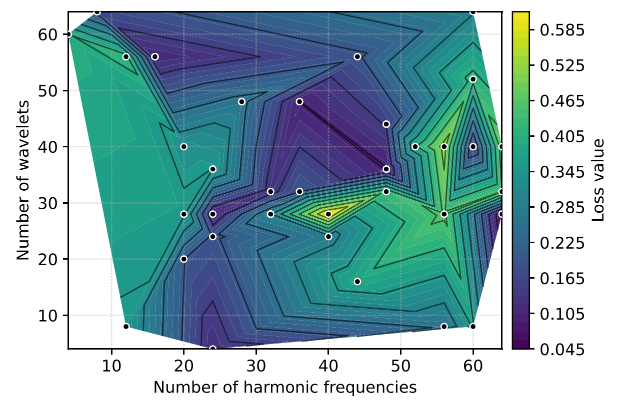

Fig. 7. Loss value (RMSE) for variation of number of wavelets and frequencies.

improves the predictive ability of the model, which is the purpose of this work. The filtered signal presented in this subsection is considered for future analysis in relation to the use of the proposed model.

5.4. Hypertuning study

The main hyperparameters of the SinWavKAN model are the number of harmonics employed and the number of wavelets, since these components constitute the fundamental principle of the model. Manually adjusting these hyperparameters may lead to the selection of inadequate or suboptimal values. For this reason, an automatic optimization procedure based on hypertuning is adopted.

During the optimization process, multiple combinations of hyperparameter configurations are iteratively evaluated until the minimum value of the defined loss function is achieved. In this work, the adopted metric is the root mean square error RMSE. The trials conducted throughout the optimization process are presented in Fig. 6. Whenever a new hyperparameter combination outperforms all previously tested configurations, the model stores these values, which are then defined as optimal.

In the optimization process, an optimal number of 28 wavelets was obtained, using 64 frequencies, with the bias term enabled. Fig. 7 presents a performance analysis based on the comparison between the number of wavelets and the number of frequencies. The results indicate that the best performance associated with the number of frequencies occurs in the range from 10 to 64. In contrast, the best values related to the number of wavelets are distributed across different configurations, suggesting a lower sensitivity of the model to this hyperparameter within the evaluated range. The optimal values obtained from this optimization process are used in the subsequent analyses presented in this paper.

Considering this optimized configuration, the Opt-EWT-SinWavKAN model is retrained and evaluated in comparison with other models. These results are presented in the next subsection. The prediction example in relation to the original signal is shown in Fig. 8, where the forecasts are 30 steps ahead using 80% of the data for training and 20% for testing. The results show that the predictions are quite promising with low variability, which is promising for use in the field.

5.5. KAN-based models analysis

To perform benchmarking, we compare the prediction horizons of the proposed Opt-EWT-SinWavKAN model with other KAN-based models in Section 5.5.1. In addition, in Section 5.5.2, the influence of data usage is analyzed through window size.

5.5.1. Horizon evaluation

Tables 4 and 5 analyze the effect of increasing the forecasting horizon from 1 up to 30. A clear and systematic degradation in performance is observed for all baseline models as the horizon increases, while the proposed method exhibits a markedly slower error growth rate. In these experiments, a window size of 5 was considered as the comparative standard.

At horizon equal to 1 (see Tables 4), the proposed model achieves an MAE of $1.96{\times}10^{-2}$, compared to $1.36{\times}10^{-1}$ for the best baseline fastKAN. This corresponds to an absolute reduction of 1.16×10^{-1} and a relative improvement of approximately 85.6%. In RMSE terms, the reduction from $9.24{\times}10^{-2}$ to $1.68{\times}10^{-2}$ represents an 81.8% improvement. SMAPE is reduced from 38.5 to 8.81, yielding a relative decrease of 77.1%.

When the horizon increases to 3, baseline MAE values rise to the range 2.07×10^{-1}-2.26×10^{-1} for fastKAN and efficientKAN, whereas the proposed model maintains an MAE of 1.95×10^{-2}. This corresponds to a reduction of approximately 90.6% relative to fastKAN. RMSE follows a similar trend, decreasing from $1.44{\times}10^{-1}$ to $1.48{\times}10^{-2}$, which is nearly a one order of magnitude reduction. SMAPE is reduced from values above 52 to 10.3, confirming strong robustness to short multi-step forecasting.

At horizon equal to 5, all baseline models exceed 2.4×10^{-1} in MAE, while the proposed approach remains at 2.33×10^{-2}. Compared to fastKAN, this represents a relative MAE reduction of 90.6%, and compared to StandardKAN, the reduction exceeds 93.6%. RMSE shows consistent behavior, with the proposed model achieving 1.79×10^{-2} versus 1.75×10^{-1} for fastKAN, corresponding to a 89.8% reduction. SMAPE is reduced from approximately 58-62 down to 10.4.

For a horizon equal to 10, the numerical gap remains substantial. The proposed model records an MAE of 2.70×10^{-2}, whereas the best baseline remains above $2.90{\times}10^{-1}$. This corresponds to an improvement of approximately 90.7%. RMSE decreases from 2.00×10^{-1} to 1.97×10^{-2}, yielding a relative reduction of 90.2%. SMAPE decreases from values close to 65-70 to 10.6, reinforcing the stability of the proposed architecture under increasing prediction horizons.

For longer horizons reported in Table 5, the relative advantage remains pronounced, although absolute errors increase for all models. At horizon equal to 15, the proposed method achieves an MAE of 3.50×10^{-2}, while the strongest baseline fastKAN reports 3.20×10^{-1}. This corresponds to an 89.1% MAE reduction. RMSE decreases from 2.31×10^{-1} to 2.48×10^{-2}, yielding an 89.3% reduction. SMAPE is reduced from approximately 71 to 13.4, corresponding to an 81.1% decrease. At horizon equal to 20, baseline MAE values cluster around 3.30×10^{-1} to 3.36×10^{-1}, whereas the proposed model maintains 3.87×10^{-2}. This yields an MAE reduction of approximately 88.4%. RMSE is reduced from 2.45×10^{-1} to 2.59×10^{-2}, corresponding to an 89.4% reduction. SMAPE decreases from about 75 to 13.8, representing an 81.6% improvement.

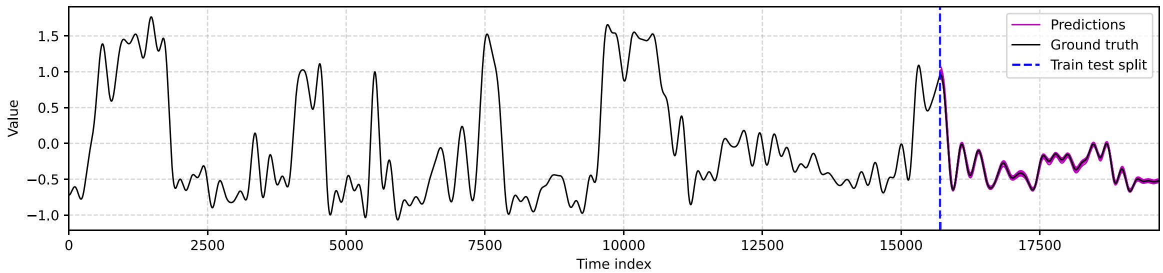

Fig. 8. Example of predictions versus ground truth for 30 steps ahead.

Table 4
Horizon analysis of KAN-based models (very short-term).

Horizon	Model	MAE	RMSE	SMAPE (%)	Params	Training time (s)	Inference time (s)
1	StandardKAN	3.56×10^{-1}	2.62×10^{-1}	8.68×10^{1}	9.70×10^{1}	9.06×10^{-1}	2.25×10^{-3}
	FastKAN	1.36×10^{-1}	9.24×10^{-2}	3.85×10^{1}	1.12×10^{4}	1.14×10^{1}	9.68×10^{-3}
	EfficientKAN	1.93×10^{-1}	1.30×10^{-1}	4.91×10^{1}	1.22×10^{4}	5.50×10^{1}	5.62×10^{-2}
	ReLUKAN	2.98×10^{-1}	2.27×10^{-1}	7.85×10^{1}	1.61×10^{2}	6.77×10^{-1}	9.37×10^{-4}
	SirenKAN	2.24×10^{-1}	1.47×10^{-1}	5.51×10^{1}	2.41×10^{2}	9.29×10^{-1}	3.26×10^{-3}
	ChebyKAN	7.36×10^{-1}	4.43×10^{-1}	1.51×10^{2}	4.10×10^{1}	6.12×10^{-1}	1.80×10^{-3}
	WavKAN	1.86×10^{-1}	1.31×10^{-1}	4.99×10^{1}	2.41×10^{2}	2.63	2.05×10^{-3}
	RBFKAN	2.56×10^{-1}	1.84×10^{-1}	6.50×10^{1}	2.41×10^{2}	1.39	1.40×10^{-3}
	FourierKAN	2.57×10^{-1}	1.78×10^{-1}	5.70×10^{1}	1.61×10^{2}	6.86×10^{-1}	4.20×10^{-3}
	Opt-EWT-SinWavKAN	1.96×10^{-2}	1.68×10^{-2}	8.81	1.38×10^{3}	1.07×10^{1}	7.47×10^{-3}
3	StandardKAN	3.69×10^{-1}	2.77×10^{-1}	9.50×10^{1}	2.59×10^{2}	3.06	6.77×10^{-3}
	fastKAN	2.07×10^{-1}	1.44×10^{-1}	5.24×10^{1}	1.17×10^{4}	1.30×10^{1}	1.06×10^{-2}
	efficientKAN	2.26×10^{-1}	1.56×10^{-1}	5.50×10^{1}	1.28×10^{4}	5.63×10^{1}	9.32×10^{-2}
	ReLUKAN	2.67×10^{-1}	2.01×10^{-1}	6.70×10^{1}	4.83×10^{2}	2.31	2.31×10^{-3}
	SirenKAN	2.56×10^{-1}	1.73×10^{-1}	6.05×10^{1}	7.23×10^{2}	3.27	2.89×10^{-3}
	ChebyKAN	8.15×10^{-1}	4.31×10^{-1}	1.12×10^{2}	1.23×10^{2}	6.85×10^{-1}	1.54×10^{-3}
	WavKAN	2.58×10^{-1}	1.85×10^{-1}	6.16×10^{1}	7.23×10^{2}	9.00	5.88×10^{-3}
	RBFKAN	2.24×10^{-1}	1.62×10^{-1}	5.62×10^{1}	7.23×10^{2}	4.68	4.55×10^{-3}
	FourierKAN	3.58×10^{-1}	2.46×10^{-1}	8.10×10^{1}	4.83×10^{2}	1.86	2.81×10^{-3}
	Opt-EWT-SinWavKAN	1.95×10^{-2}	1.48×10^{-2}	1.03×10^{1}	4.14×10^{3}	3.96×10^{1}	2.30×10^{-2}
5	StandardKAN	3.65×10^{-1}	2.78×10^{-1}	9.70×10^{1}	4.21×10^{2}	2.36	1.93×10^{-3}
	fastKAN	2.49×10^{-1}	1.75×10^{-1}	6.20×10^{1}	1.23×10^{4}	1.15×10^{1}	1.06×10^{-2}
	efficientKAN	2.53×10^{-1}	1.76×10^{-1}	5.96×10^{1}	1.34×10^{4}	5.56×10^{1}	6.19×10^{-2}
	ReLUKAN	3.25×10^{-1}	2.49×10^{-1}	8.03×10^{1}	8.05×10^{2}	3.87	4.03×10^{-3}
	SirenKAN	2.61×10^{-1}	1.77×10^{-1}	6.05×10^{1}	1.20×10^{3}	5.83	5.89×10^{-3}
	ChebyKAN	8.40×10^{-1}	4.49×10^{-1}	1.11×10^{2}	2.05×10^{2}	1.18	1.71×10^{-3}
	WavKAN	2.44×10^{-1}	1.66×10^{-1}	5.81×10^{1}	1.20×10^{3}	1.48×10^{1}	9.57×10^{-3}
	RBFKAN	2.94×10^{-1}	2.09×10^{-1}	6.62×10^{1}	1.20×10^{3}	8.00	6.20×10^{-3}
	FourierKAN	3.56×10^{-1}	2.44×10^{-1}	7.63×10^{1}	8.05×10^{2}	2.90	3.84×10^{-3}
	Opt-EWT-SinWavKAN	2.33×10^{-2}	1.79×10^{-2}	1.04×10^{1}	6.90×10^{3}	1.09×10^{2}	3.68×10^{-2}
10	StandardKAN	4.23×10^{-1}	3.36×10^{-1}	1.30×10^{2}	8.26×10^{2}	4.54	3.30×10^{-3}
	fastKAN	2.96×10^{-1}	2.11×10^{-1}	6.93×10^{1}	1.38×10^{4}	1.21×10^{1}	1.04×10^{-2}
	efficientKAN	3.02×10^{-1}	2.16×10^{-1}	6.82×10^{1}	1.50×10^{4}	5.59×10^{1}	9.97×10^{-2}
	ReLUKAN	3.41×10^{-1}	2.59×10^{-1}	8.40×10^{1}	1.61×10^{3}	8.29	8.06×10^{-3}
	SirenKAN	3.11×10^{-1}	2.15×10^{-1}	6.94×10^{1}	2.41×10^{3}	1.17×10^{1}	9.40×10^{-3}
	ChebyKAN	7.23×10^{-1}	4.23×10^{-1}	1.18×10^{2}	4.10×10^{2}	1.40	2.67×10^{-3}
	WavKAN	3.03×10^{-1}	2.17×10^{-1}	6.76×10^{1}	2.41×10^{3}	3.06×10^{1}	1.96×10^{-2}
	RBFKAN	3.18×10^{-1}	2.30×10^{-1}	7.01×10^{1}	2.41×10^{3}	1.62×10^{1}	1.09×10^{-2}
	FourierKAN	3.69×10^{-1}	2.59×10^{-1}	7.87×10^{1}	1.61×10^{3}	5.14	6.71×10^{-3}
	Opt-EWT-SinWavKAN	2.70×10^{-2}	1.97×10^{-2}	1.06×10^{1}	1.38×10^{4}	2.59×10^{2}	8.85×10^{-2}

For the most challenging cases, horizons equal to 25 and 30, the proposed model continues to outperform all competitors by large margins. At horizon equal to 25, the MAE reduction relative to fastKAN is approximately 85.3%, while at horizon equal to 30 it remains close to 84.3%. RMSE reductions remain above 85% in both cases. SMAPE, which exceeds 75 for all baseline methods, is constrained below 18.2 by the proposed approach.

The numerical evidence across all horizons demonstrates that the proposed method consistently achieves error reductions between 80% and 93% relative to the strongest KAN-based baselines. Most importantly, the slope of error growth with respect to the prediction horizon is significantly smaller, indicating superior long-range temporal generalization. Although this performance gain is accompanied by increased parameter counts and training time for large horizons, the magnitude of the accuracy improvements is disproportionally large, confirming that the proposed architecture delivers a fundamentally different accuracy scaling behavior rather than a marginal refinement of existing KAN variants.

5.5.2. Window evaluation

Table 6 reports the forecasting performance considering different window sizes and multiple KAN-based architectures. The results highlight clear and consistent trends across all evaluation metrics. For Window equal to 5, the proposed model achieves an MAE of 1.78×10^{-2}, whereas the best competing baseline fastKAN reports 1.36×10^{-1}. This

Table 5
Horizon analysis of KAN-based models (short-term).

Horizon	Model	MAE	RMSE	SMAPE (%)	Params	Training time (s)	Inference time (s)
15	StandardKAN	4.22×10^{-1}	3.34×10^{-1}	1.26×10^{2}	1.23×10^{3}	6.13	4.33×10^{-3}
	fastKAN	3.20×10^{-1}	2.31×10^{-1}	7.15×10^{1}	1.52×10^{4}	1.18×10^{1}	1.09×10^{-2}
	efficientKAN	3.27×10^{-1}	2.40×10^{-1}	7.27×10^{1}	1.66×10^{4}	5.52×10^{1}	9.32×10^{-2}
	ReLUKAN	3.31×10^{-1}	2.47×10^{-1}	7.64×10^{1}	2.42×10^{3}	1.33×10^{1}	1.43×10^{-2}
	SirenKAN	3.30×10^{-1}	2.33×10^{-1}	7.35×10^{1}	3.62×10^{3}	1.98×10^{1}	1.82×10^{-2}
	ChebyKAN	9.92×10^{-1}	5.51×10^{-1}	1.35×10^{2}	6.15×10^{2}	2.35	3.22×10^{-3}
	WavKAN	3.35×10^{-1}	2.45×10^{-1}	7.30×10^{1}	3.62×10^{3}	9.45×10^{1}	3.30×10^{-2}
	RBFKAN	3.51×10^{-1}	2.57×10^{-1}	7.66×10^{1}	3.62×10^{3}	4.63×10^{1}	1.79×10^{-2}
	FourierKAN	3.98×10^{-1}	2.85×10^{-1}	8.23×10^{1}	2.42×10^{3}	7.66	1.01×10^{-2}
	Opt-EWT-SinWavKAN	3.50×10^{-2}	2.48×10^{-2}	1.34×10^{1}	2.07×10^{4}	3.87×10^{2}	1.28×10^{-1}
20	StandardKAN	4.05×10^{-1}	3.17×10^{-1}	1.14×10^{2}	1.64×10^{3}	7.62	5.86×10^{-3}
	fastKAN	3.34×10^{-1}	2.45×10^{-1}	7.49×10^{1}	1.67×10^{4}	1.19×10^{1}	1.17×10^{-2}
	efficientKAN	3.36×10^{-1}	2.49×10^{-1}	7.40×10^{1}	1.82×10^{4}	5.57×10^{1}	6.64×10^{-2}
	ReLUKAN	3.55×10^{-1}	2.69×10^{-1}	8.65×10^{1}	3.22×10^{3}	2.27×10^{1}	1.63×10^{-2}
	SirenKAN	3.52×10^{-1}	2.55×10^{-1}	7.85×10^{1}	4.82×10^{3}	3.16×10^{1}	2.07×10^{-2}
	ChebyKAN	8.37×10^{-1}	4.59×10^{-1}	1.24×10^{2}	8.20×10^{2}	3.40	5.55×10^{-3}
	WavKAN	3.46×10^{-1}	2.55×10^{-1}	7.47×10^{1}	4.82×10^{3}	1.48×10^{2}	4.25×10^{-2}
	RBFKAN	3.59×10^{-1}	2.66×10^{-1}	7.80×10^{1}	4.82×10^{3}	6.79×10^{1}	3.06×10^{-2}
	FourierKAN	3.95×10^{-1}	2.89×10^{-1}	8.31×10^{1}	3.22×10^{3}	1.17×10^{1}	1.11×10^{-2}
	Opt-EWT-SinWavKAN	3.87×10^{-2}	2.59×10^{-2}	1.38×10^{1}	2.76×10^{4}	5.29×10^{2}	2.28×10^{-1}
25	StandardKAN	4.08×10^{-1}	3.19×10^{-1}	1.14×10^{2}	2.04×10^{3}	9.63	6.99×10^{-3}
	fastKAN	3.48×10^{-1}	2.59×10^{-1}	7.73×10^{1}	1.81×10^{4}	1.20×10^{1}	1.06×10^{-2}
	efficientKAN	3.45×10^{-1}	2.59×10^{-1}	7.59×10^{1}	1.98×10^{4}	5.61×10^{1}	5.99×10^{-2}
	ReLUKAN	3.70×10^{-1}	2.81×10^{-1}	8.91×10^{1}	4.02×10^{3}	5.16×10^{1}	1.97×10^{-2}
	SirenKAN	3.62×10^{-1}	2.67×10^{-1}	8.13×10^{1}	6.02×10^{3}	6.21×10^{1}	2.51×10^{-2}
	ChebyKAN	9.46×10^{-1}	5.17×10^{-1}	1.34×10^{2}	1.02×10^{3}	3.63	4.84×10^{-3}
	WavKAN	3.51×10^{-1}	2.61×10^{-1}	7.60×10^{1}	6.02×10^{3}	2.01×10^{2}	5.95×10^{-2}
	RBFKAN	3.60×10^{-1}	2.69×10^{-1}	7.84×10^{1}	6.02×10^{3}	1.07×10^{2}	3.36×10^{-2}
	FourierKAN	3.96×10^{-1}	2.90×10^{-1}	8.26×10^{1}	4.02×10^{3}	3.29×10^{1}	1.41×10^{-2}
	Opt-EWT-SinWavKAN	5.12×10^{-2}	3.43×10^{-2}	1.78×10^{1}	3.45×10^{4}	6.59×10^{2}	4.23×10^{-1}
30	StandardKAN	4.04×10^{-1}	3.17×10^{-1}	1.12×10^{2}	2.45×10^{3}	2.87×10^{1}	8.16×10^{-3}
	fastKAN	3.54×10^{-1}	2.64×10^{-1}	7.89×10^{1}	1.96×10^{4}	1.18×10^{1}	1.00×10^{-2}
	efficientKAN	3.48×10^{-1}	2.62×10^{-1}	7.63×10^{1}	2.14×10^{4}	5.58×10^{1}	5.95×10^{-2}
	ReLUKAN	3.66×10^{-1}	2.78×10^{-1}	8.81×10^{1}	4.83×10^{3}	3.32×10^{1}	3.03×10^{-2}
	SirenKAN	3.64×10^{-1}	2.69×10^{-1}	8.17×10^{1}	7.23×10^{3}	6.27×10^{1}	3.44×10^{-2}
	ChebyKAN	8.92×10^{-1}	5.17×10^{-1}	1.35×10^{2}	1.23×10^{3}	4.61	5.35×10^{-3}
	WavKAN	3.59×10^{-1}	2.68×10^{-1}	7.71×10^{1}	7.23×10^{3}	2.17×10^{2}	6.94×10^{-2}
	RBFKAN	3.64×10^{-1}	2.72×10^{-1}	7.92×10^{1}	7.23×10^{3}	1.09×10^{2}	4.05×10^{-2}
	FourierKAN	4.03×10^{-1}	2.98×10^{-1}	8.41×10^{1}	4.83×10^{3}	1.62×10^{1}	1.67×10^{-2}
	Opt-EWT-SinWavKAN	5.57×10^{-2}	3.65×10^{-2}	1.82×10^{1}	4.14×10^{4}	7.89×10^{2}	4.06×10^{-1}

corresponds to an absolute reduction of 1.18×10^{-1} and a relative improvement of approximately 86.9% in MAE. A similar trend is observed for RMSE, which decreases from 9.24×10^{-2} with fastKAN to 1.30×10^{-2}, yielding a relative reduction close to 85.9%. In terms of SMAPE, the proposed approach reaches 8.35, while the closest competitor remains above 38, representing a reduction of nearly 78%. These gains are achieved with 1.38×10^{3} parameters, which is roughly one order of magnitude about eight times smaller than fastKAN.

For a window equal to 10, the numerical advantage remains strong. The proposed model attains an MAE of 1.74×10^{-2} compared to 1.76×10^{-1} for WavKAN. This corresponds to an approximate tenfold reduction in MAE. The RMSE drops from values around 1.15×10^{-1} for WavKAN to 7.98×10^{-3}, which represents a relative decrease of about 92.9%. SMAPE is reduced from values in the range of 44 to 46 down to 6.40, confirming a consistent and substantial numerical improvement.

When the window size increases to 25, the overall error levels rise for all models, but the proposed method still delivers the lowest metrics. Its MAE of 1.36×10^{-1} is lower than the best competing baseline SirenKAN at 1.66×10^{-1}, corresponding to a relative improvement of approximately 18.1%. For RMSE, the reduction is from 1.19×10^{-1} to 9.64×10^{-2}, which corresponds to an improvement of about 19.0%. SMAPE decreases from roughly 47 to 42.9, yielding a reduction close to 9.5%.

For the largest configuration, a window equal to 50, the contrast becomes even clearer. This corresponds to an absolute reduction of 1.59×10^{-1} and a relative improvement of approximately 85.0%. The RMSE shows a similar behavior, decreasing from 1.25×10^{-1} to 2.08×10^{-2}, which represents an 83.4% reduction. In terms of SMAPE, the proposed approach reaches 16.2, whereas most baselines remain between 49 and 90, indicating reductions that range from about 67% to over 80%. Although the parameter count increases to 1.38×10^{4} and the training time rises accordingly, the numerical accuracy gains remain disproportionally large when compared to the additional computational cost.

A quantitative comparison across all window sizes shows that the proposed method consistently reduces MAE and RMSE by factors ranging from about 5 to more than 10 relative to the best competing KAN variants for small and large windows, while still maintaining noticeable improvements for intermediate window sizes. These numerical margins clearly demonstrate that the performance gains are not marginal but substantial and systematic across all evaluated configurations.

5.6. Benchmarking and generalization analysis

The benchmarking analysis evaluates the predictive capability and computational efficiency of the proposed model across three distinct hydrological datasets, namely `FUNDAO` (Table 7), `SEGREDO` (Table 8), and `FOZ DO CHAPECÓ` (Table 9). The comparison includes a wide range of state of the art deep learning forecasting architectures representing different modeling paradigms, including MLP-based, convolutional, recurrent, and transformer-based approaches. The evaluation metrics comprise RMSE, MAE, and SMAPE, which quantify forecasting accuracy, while training and inference times measure computational performance.

Table 6
Window size analysis of KAN-based models.

Window	Model	MAE	RMSE	SMAPE (%)	Params	Training time (s)	Inference time (s)
5	StandardKAN	3.83×10^{-1}	2.99×10^{-1}	1.10×10^{2}	9.70×10^{1}	9.49×10^{-1}	8.88×10^{-4}
	fastKAN	1.36×10^{-1}	9.24×10^{-2}	3.85×10^{1}	1.12×10^{4}	1.09×10^{1}	1.18×10^{-2}
	efficientKAN	1.93×10^{-1}	1.30×10^{-1}	4.91×10^{1}	1.22×10^{4}	5.52×10^{1}	5.55×10^{-2}
	ReLUKAN	2.98×10^{-1}	2.27×10^{-1}	7.85×10^{1}	1.61×10^{2}	6.26×10^{-1}	8.88×10^{-4}
	SirenKAN	2.24×10^{-1}	1.47×10^{-1}	5.51×10^{1}	2.41×10^{2}	8.43×10^{-1}	3.12×10^{-3}
	ChebyKAN	7.36×10^{-1}	4.43×10^{-1}	1.51×10^{2}	4.10×10^{1}	4.13×10^{-1}	1.05×10^{-3}
	WavKAN	1.86×10^{-1}	1.31×10^{-1}	4.99×10^{1}	2.41×10^{2}	2.36	1.97×10^{-3}
	RBFKAN	2.56×10^{-1}	1.84×10^{-1}	6.50×10^{1}	2.41×10^{2}	1.47	1.45×10^{-3}
	FourierKAN	2.57×10^{-1}	1.78×10^{-1}	5.70×10^{1}	1.61×10^{2}	8.52×10^{-1}	5.24×10^{-3}
	Opt-EWT-SinWavKAN	1.78×10^{-2}	1.30×10^{-2}	8.35	1.38×10^{3}	1.25×10^{1}	6.85×10^{-3}
10	StandardKAN	2.91×10^{-1}	2.13×10^{-1}	7.40×10^{1}	1.77×10^{2}	1.33	1.52×10^{-3}
	fastKAN	2.17×10^{-1}	1.63×10^{-1}	5.75×10^{1}	1.26×10^{4}	1.22×10^{1}	1.13×10^{-2}
	efficientKAN	2.02×10^{-1}	1.39×10^{-1}	5.14×10^{1}	1.38×10^{4}	5.76×10^{1}	6.24×10^{-2}
	ReLUKAN	2.60×10^{-1}	2.01×10^{-1}	6.63×10^{1}	3.21×10^{2}	1.27	1.45×10^{-3}
	SirenKAN	1.83×10^{-1}	1.21×10^{-1}	4.46×10^{1}	4.81×10^{2}	1.95	2.11×10^{-3}
	ChebyKAN	1.06	3.82×10^{-1}	1.01×10^{2}	8.10×10^{1}	7.86×10^{-1}	1.91×10^{-3}
	WavKAN	1.76×10^{-1}	1.15×10^{-1}	4.56×10^{1}	4.81×10^{2}	5.50	5.28×10^{-3}
	RBFKAN	1.97×10^{-1}	1.47×10^{-1}	5.22×10^{1}	4.81×10^{2}	3.10	2.48×10^{-3}
	FourierKAN	4.72×10^{-1}	3.23×10^{-1}	8.76×10^{1}	3.21×10^{2}	1.61	2.74×10^{-3}
	Opt-EWT-SinWavKAN	1.74×10^{-2}	7.98×10^{-3}	6.40	2.76×10^{3}	2.72×10^{1}	1.50×10^{-2}
25	StandardKAN	2.76×10^{-1}	2.12×10^{-1}	6.95×10^{1}	4.17×10^{2}	3.75	3.21×10^{-3}
	fastKAN	2.28×10^{-1}	1.75×10^{-1}	6.20×10^{1}	1.70×10^{4}	1.46×10^{1}	1.82×10^{-2}
	efficientKAN	2.16×10^{-1}	1.53×10^{-1}	5.60×10^{1}	1.86×10^{4}	6.29×10^{1}	7.58×10^{-2}
	ReLUKAN	3.26×10^{-1}	2.48×10^{-1}	7.62×10^{1}	8.01×10^{2}	3.92	4.67×10^{-3}
	SirenKAN	1.66×10^{-1}	1.19×10^{-1}	4.74×10^{1}	1.20×10^{3}	4.77	4.23×10^{-3}
	ChebyKAN	1.68	8.79×10^{-1}	1.52×10^{2}	2.01×10^{2}	2.21	4.59×10^{-3}
	WavKAN	2.14×10^{-1}	1.52×10^{-1}	5.48×10^{1}	1.20×10^{3}	1.46×10^{1}	9.01×10^{-3}
	RBFKAN	2.15×10^{-1}	1.48×10^{-1}	5.58×10^{1}	1.20×10^{3}	7.72	5.08×10^{-3}
	FourierKAN	5.69×10^{-1}	4.16×10^{-1}	9.88×10^{1}	8.01×10^{2}	4.32	5.96×10^{-3}
	Opt-EWT-SinWavKAN	1.36×10^{-1}	9.64×10^{-2}	4.29×10^{1}	6.90×10^{3}	1.23×10^{2}	4.15×10^{-2}
50	StandardKAN	3.57×10^{-1}	2.72×10^{-1}	9.05×10^{1}	8.17×10^{2}	7.87	6.53×10^{-3}
	fastKAN	2.67×10^{-1}	2.07×10^{-1}	6.88×10^{1}	2.42×10^{4}	1.87×10^{1}	1.64×10^{-2}
	efficientKAN	2.44×10^{-1}	1.77×10^{-1}	6.18×10^{1}	2.66×10^{4}	7.29×10^{1}	1.00×10^{-1}
	ReLUKAN	2.71×10^{-1}	2.03×10^{-1}	6.47×10^{1}	1.60×10^{3}	7.88	6.88×10^{-3}
	SirenKAN	3.94×10^{-1}	2.97×10^{-1}	8.59×10^{1}	2.40×10^{3}	1.07×10^{1}	8.71×10^{-3}
	ChebyKAN	2.76	1.44	1.54×10^{2}	4.01×10^{2}	4.77	8.92×10^{-3}
	WavKAN	2.38×10^{-1}	1.65×10^{-1}	6.02×10^{1}	2.40×10^{3}	4.43×10^{1}	2.20×10^{-2}
	RBFKAN	4.01×10^{-1}	3.03×10^{-1}	8.70×10^{1}	2.40×10^{3}	1.57×10^{1}	1.11×10^{-2}
	FourierKAN	7.74×10^{-1}	5.86×10^{-1}	1.17×10^{2}	1.60×10^{3}	8.89	1.35×10^{-2}
	Opt-EWT-SinWavKAN	2.80×10^{-2}	2.08×10^{-2}	1.62×10^{1}	1.38×10^{4}	2.49×10^{2}	7.52×10^{-2}

Table 7
Comparative analysis for `FUNDAO` dataset.

Model	RMSE	MAE	SMAPE (%)	Training time (s)	Inference time (s)
TSMixer	6.20×10^{-1}	5.49×10^{-1}	1.02×10^{2}	7.58	4.45×10^{-1}
iTransformer	4.75×10^{-1}	3.96×10^{-1}	9.07×10^{1}	1.20×10^{1}	4.04×10^{-1}
TSMixerx	3.12×10^{-1}	2.72×10^{-1}	8.25×10^{1}	9.05	1.97×10^{-1}
TFT	3.39×10^{-1}	2.94×10^{-1}	9.91×10^{1}	7.02	1.86×10^{-1}
VanillaRNN	7.05×10^{-1}	6.44×10^{-1}	1.09×10^{2}	3.55	1.48×10^{-1}
DilatedRNN	6.76×10^{-1}	6.35×10^{-1}	1.11×10^{2}	5.47	1.79×10^{-1}
NHITS	3.25×10^{-1}	2.80×10^{-1}	8.27×10^{1}	4.10	1.69×10^{-1}
TCN	2.93×10^{-1}	2.56×10^{-1}	8.08×10^{1}	3.98	2.15×10^{-1}
BiTCN	5.99×10^{-1}	5.25×10^{-1}	1.01×10^{2}	4.09	1.49×10^{-1}
LSTM	7.30×10^{-1}	6.99×10^{-1}	1.15×10^{2}	4.89	1.42×10^{-1}
NBEATS	3.04×10^{-1}	2.79×10^{-1}	8.60×10^{1}	4.32	1.66×10^{-1}
NBEATSx	3.04×10^{-1}	2.79×10^{-1}	8.60×10^{1}	4.14	2.24×10^{-1}
GRU	7.22×10^{-1}	6.86×10^{-1}	1.14×10^{2}	4.12	1.45×10^{-1}
Informer	6.02×10^{-1}	5.19×10^{-1}	9.98×10^{1}	7.59	1.84×10^{-1}
TiDE	3.82×10^{-1}	3.21×10^{-1}	1.07×10^{2}	4.06	1.66×10^{-1}
PatchTST	3.36×10^{-1}	2.88×10^{-1}	1.08×10^{2}	5.90	1.52×10^{-1}
FEDformer	4.71×10^{-1}	3.91×10^{-1}	9.02×10^{1}	6.91	1.67×10^{-1}
DeepAR	8.21×10^{-1}	8.01×10^{-1}	1.21×10^{2}	4.20	2.04×10^{-1}
TimesNet	3.30×10^{-1}	3.09×10^{-1}	9.82×10^{1}	1.59×10^{1}	1.83×10^{-1}
SinWaveKAN	2.70×10^{-1}	1.68×10^{-1}	5.74×10^{1}	1.58×10^{2}	1.32×10^{-1}
Opt-EWT-SinWavKAN	2.40×10^{-2}	1.82×10^{-2}	9.39	1.29×10^{2}	7.56×10^{-2}

For the `FUNDAO` dataset (results presented in Table 7), the proposed model demonstrates a substantial performance improvement compared to all baseline methods. The obtained RMSE of 2.40×10^{-2} and MAE of 1.82×10^{-2} are approximately one order of magnitude lower than the

Table 8
Comparative analysis for SEGREDO dataset.

Model	RMSE	MAE	SMAPE (%)	Training time (s)	Inference time (s)
TSMixer	2.15×10^{-1}	2.12×10^{-1}	3.31×10^{1}	1.97×10^{1}	1.56×10^{-1}
iTransformer	1.23×10^{-1}	1.21×10^{-1}	2.03×10^{1}	5.91	2.10×10^{-1}
TSMixerx	1.35×10^{-1}	1.34×10^{-1}	2.23×10^{1}	4.14	1.76×10^{-1}
TFT	1.59×10^{-1}	1.58×10^{-1}	2.58×10^{1}	7.59	1.48×10^{-1}
VanillaRNN	2.16×10^{-1}	2.15×10^{-1}	3.36×10^{1}	3.30	1.58×10^{-1}
DilatedRNN	2.16×10^{-1}	2.15×10^{-1}	3.35×10^{1}	5.53	1.28×10^{-1}
NHITS	1.16×10^{-1}	1.15×10^{-1}	1.94×10^{1}	4.04	1.44×10^{-1}
TCN	1.57×10^{-1}	1.54×10^{-1}	2.53×10^{1}	3.48	1.92×10^{-1}
BiTCN	1.98×10^{-1}	1.97×10^{-1}	3.12×10^{1}	3.69	1.51×10^{-1}
LSTM	2.14×10^{-1}	2.13×10^{-1}	3.34×10^{1}	3.76	1.35×10^{-1}
NBEATS	1.06×10^{-1}	1.05×10^{-1}	1.79×10^{1}	3.95	1.63×10^{-1}
NBEATSx	1.06×10^{-1}	1.05×10^{-1}	1.79×10^{1}	4.10	1.71×10^{-1}
GRU	2.16×10^{-1}	2.16×10^{-1}	3.37×10^{1}	3.51	1.53×10^{-1}
Informer	2.02×10^{-1}	1.97×10^{-1}	3.11×10^{1}	8.45	1.91×10^{-1}
TiDE	2.32×10^{-1}	2.14×10^{-1}	4.19×10^{1}	3.78	1.28×10^{-1}
PatchTST	8.32×10^{-2}	7.81×10^{-2}	1.36×10^{1}	5.81	1.76×10^{-1}
FEDformer	1.65×10^{-1}	1.62×10^{-1}	2.64×10^{1}	1.11×10^{1}	2.45×10^{-1}
DeepAR	8.27×10^{-1}	7.47×10^{-1}	7.75×10^{1}	6.09	4.27×10^{-1}
TimesNet	1.07×10^{-1}	1.03×10^{-1}	1.76×10^{1}	1.64×10^{1}	3.71×10^{-1}
SinWaveKAN	1.23×10^{-1}	9.52×10^{-2}	2.32×10^{1}	1.34×10^{2}	8.11×10^{-2}
Opt-EWT-SinWavKAN	6.40×10^{-2}	4.36×10^{-2}	9.98	1.35×10^{2}	7.20×10^{-2}

second-best-performing models, such as TCN, NBEATS, and TSMixerx, which present RMSE values around 3.0×10^{-1}. Similarly, the SMAPE value of 9.39 significantly outperforms competing approaches, whose SMAPE values remain above 80. These results indicate that the proposed architecture captures complex nonlinear temporal dynamics with considerably higher precision. Although the training time of 1.29×10^{2} seconds is larger than most baseline models, the inference time of 7.56×10^{-2} seconds remains competitive and suitable for real-time applications. The results suggest that the additional training complexity enables the model to learn more expressive temporal representations, resulting in superior forecasting accuracy.

The standalone SinWaveKAN model, without the EWT preprocessing stage, demonstrates competitive predictive performance across the three hydropower datasets, particularly considering its relatively simple architectural design based purely on sinusoidal and wavelet-inspired basis learning. For the FUNDAO dataset, SinWaveKAN achieves the best overall performance among all baseline models, reaching an RMSE of 2.70×10^{-1}, MAE of 1.68×10^{-1}, and SMAPE of 5.74×10^{1}%, clearly outperforming strong competitors such as TCN, NHITS, and NBEATS.

For the SEGREDO dataset (see Table 8), the proposed model again achieves the best overall performance across all error metrics. The RMSE and MAE values of 6.40×10^{-2} and 4.36×10^{-2}, respectively, outperform strong transformer-based and patch-based models, including PatchTST and NBEATS, which represent the strongest baselines for this dataset. The SMAPE value of 9.98 further confirms the robustness of the proposed model, demonstrating superior relative error performance under varying magnitude scales. The training time of approximately 1.35×10^{2} seconds is comparable to other high-capacity architectures, while the inference time of 7.20×10^{-2} seconds is among the fastest methods, indicating that the model maintains efficient forward propagation despite its increased representational power.

For the SEGREDO dataset, SinWaveKAN remains competitive, obtaining an RMSE of 1.23×10^{-1} and MAE of 9.52×10^{-2}, although slightly inferior to PatchTST and NBEATS, while still maintaining lower inference time compared to most transformer-based approaches. The consistency of these results across both FUNDAO and SEGREDO datasets highlights the strong generalization capability of the proposed framework.

For the FOZ DO CHAPECÓ dataset (see Table 9), which exhibits different temporal patterns and statistical characteristics, the proposed model continues to outperform all comparison methods in terms of RMSE and MAE, achieving values of 4.83×10^{-2} and 3.84×10^{-2} respectively. The SMAPE value of 2.06×10^{1} remains competitive with the best baseline models, demonstrating reliable relative forecasting performance. Although certain models, such as NHITS and NBEATS, achieve competitive SMAPE values, their absolute error metrics remain significantly higher than those obtained by the proposed method. The training time remains consistent with the previous datasets, while the inference time of 5.09×10^{-2} seconds represents the fastest among all evaluated methods, reinforcing the practical applicability of the model for operational forecasting environments.

In the more challenging FOZ DO CHAPECO dataset, the standalone SinWaveKAN model shows reduced accuracy (RMSE of 3.66×10^{-1}), suggesting that highly complex temporal dynamics may require additional preprocessing or decomposition strategies. These results indicate that even without signal decomposition, SinWaveKAN is capable of achieving strong forecasting performance, validating the effectiveness of sinusoidal and wavelet-inspired functional representations for learning temporal patterns.

From a global perspective, the results demonstrate that recurrent architectures such as RNN, GRU, and LSTM generally present inferior performance across all datasets, indicating limitations in capturing long term temporal dependencies and complex nonlinear interactions present in hydrological time series. Transformer-based models, including iTransformer, Informer, FEDformer, and PatchTST, provide improved accuracy but remain less competitive than the proposed approach, suggesting that attention mechanisms alone may not fully capture the multi-scale temporal structures inherent in reservoir data. Similarly, convolutional and MLP-based models such as TCN, NHITS, NBEATS, and TSMixer variants show strong baseline performance but consistently fall short of the predictive accuracy achieved by the proposed architecture.

The results show that the complementarity between sinusoidal and wavelet components arises from their different roles in time–frequency representation. Sinusoidal bases are effective for modeling global low-frequency periodic patterns, while wavelet bases capture localized high-frequency and transient variations. This creates a functional decomposition in which each basis represents different signal characteristics, reducing redundancy and improving approximation efficiency. Within SinWavKAN, this leads to a more expressive representation where sinusoidal components model the global structure and wavelets act as localized corrections.

Table 9
Comparative analysis for FOZ DO CHAPECO dataset.

Model	RMSE	MAE	SMAPE (%)	Training time (s)	Inference time (s)
TSMixer	5.56×10^{-1}	5.34×10^{-1}	4.19×10^{1}	7.89	1.82
iTransformer	2.86×10^{-1}	2.21×10^{-1}	2.04×10^{1}	9.54	2.26×10^{-1}
TSMixerx	3.13×10^{-1}	2.56×10^{-1}	2.31×10^{1}	7.94	1.44×10^{-1}
TFT	2.53×10^{-1}	2.09×10^{-1}	1.95×10^{1}	6.37	1.95×10^{-1}
VanillaRNN	3.26×10^{-1}	2.77×10^{-1}	2.45×10^{1}	3.60	1.39×10^{-1}
DilatedRNN	4.23×10^{-1}	3.79×10^{-1}	3.19×10^{1}	5.76	1.42×10^{-1}
NHITS	2.13×10^{-1}	1.71×10^{-1}	1.64×10^{1}	4.23	2.03×10^{-1}
TCN	3.49×10^{-1}	2.81×10^{-1}	2.47×10^{1}	3.89	1.30×10^{-1}
BiTCN	4.54×10^{-1}	4.23×10^{-1}	3.50×10^{1}	4.38	1.97×10^{-1}
LSTM	3.46×10^{-1}	3.33×10^{-1}	2.86×10^{1}	3.96	1.57×10^{-1}
NBEATS	2.01×10^{-1}	1.63×10^{-1}	1.57×10^{1}	3.86	1.67×10^{-1}
NBEATSx	2.01×10^{-1}	1.63×10^{-1}	1.57×10^{1}	3.78	1.93×10^{-1}
GRU	2.53×10^{-1}	2.40×10^{-1}	2.16×10^{1}	3.51	1.82×10^{-1}
Informer	3.45×10^{-1}	3.12×10^{-1}	2.72×10^{1}	8.01	1.67×10^{-1}
TiDE	7.06×10^{-1}	5.01×10^{-1}	6.13×10^{1}	3.46	1.87×10^{-1}
PatchTST	2.85×10^{-1}	2.39×10^{-1}	2.19×10^{1}	5.27	1.91×10^{-1}
FEDformer	3.32×10^{-1}	2.67×10^{-1}	2.26×10^{1}	6.61	1.68×10^{-1}
DeepAR	7.23×10^{-1}	6.83×10^{-1}	5.02×10^{1}	4.47	2.16×10^{-1}
TimesNet	2.26×10^{-1}	1.78×10^{-1}	1.71×10^{1}	1.64×10^{1}	2.32×10^{-1}
SinWaveKAN	3.66×10^{-1}	2.60×10^{-1}	6.93×10^{1}	1.46×10^{2}	6.08×10^{-2}
Opt-EWT-SinWavKAN	4.83×10^{-2}	3.84×10^{-2}	2.06×10^{1}	1.36×10^{2}	5.09×10^{-2}

Table 10
Wilcoxon signed-rank test against the best mean model for RMSE.

Compared model	Avg. Rank	Mean	p-value	Holm p-value	Significant
Opt-EWT-SinWavKAN	1	2.26×10^{-1}	–	–	–
TCN	2	2.79×10^{-1}	1.78×10^{-15}	3.73×10^{-14}	Yes
MLP	3	2.97×10^{-1}	1.54×10^{-12}	2.61×10^{-11}	Yes
NBEATS	4	3.04×10^{-1}	1.54×10^{-12}	2.61×10^{-11}	Yes
NBEATSx	5	3.04×10^{-1}	1.54×10^{-12}	2.61×10^{-11}	Yes
TSMixerx	6	3.12×10^{-1}	1.54×10^{-12}	2.61×10^{-11}	Yes
NHITS	7	3.25×10^{-1}	1.54×10^{-12}	2.61×10^{-11}	Yes
TimesNet	8	3.29×10^{-1}	1.78×10^{-15}	3.73×10^{-14}	Yes
PatchTST	9	3.36×10^{-1}	1.54×10^{-12}	2.61×10^{-11}	Yes
TFT	10	3.40×10^{-1}	1.54×10^{-12}	2.61×10^{-11}	Yes
TiDE	11	3.82×10^{-1}	1.54×10^{-12}	2.61×10^{-11}	Yes
iTransformer	12	4.48×10^{-1}	1.54×10^{-12}	2.61×10^{-11}	Yes
FEDformer	13	4.71×10^{-1}	1.78×10^{-15}	3.73×10^{-14}	Yes
BiTCN	14	5.99×10^{-1}	7.54×10^{-10}	1.51×10^{-9}	Yes
Informer	15	6.03×10^{-1}	1.78×10^{-15}	3.73×10^{-14}	Yes
TSMixer	16	6.20×10^{-1}	1.54×10^{-12}	2.61×10^{-11}	Yes
DilatedRNN	17	6.76×10^{-1}	1.54×10^{-12}	2.61×10^{-11}	Yes
RNN	18	6.99×10^{-1}	1.54×10^{-12}	2.61×10^{-11}	Yes
GRU	19	7.15×10^{-1}	1.54×10^{-12}	2.61×10^{-11}	Yes
LSTM	20	7.28×10^{-1}	1.54×10^{-12}	2.61×10^{-11}	Yes
DeepAR	21	8.21×10^{-1}	1.54×10^{-12}	2.61×10^{-11}	Yes

The benchmarking results confirm that the proposed model provides superior forecasting accuracy, strong generalization across heterogeneous hydrological datasets, and efficient inference performance. The consistent improvements across multiple evaluation metrics and datasets indicate that the model effectively captures both local and global temporal dependencies while maintaining practical computational efficiency. These findings highlight the robustness and scalability of the proposed architecture for real-world hydrological forecasting applications.

5.7. Statistical assessment

Tables 10, 11, and 12 present the results of the Wilcoxon signed rank test performed to assess the statistical significance of the performance differences between Opt-EWT-SinWavKAN and the competing forecasting models for RMSE, MAE, and SMAPE, respectively. The analysis shows that the proposed model achieves the best average rank and the lowest mean error values across all evaluated metrics.

As shown in Table 10, Opt-EWT-SinWavKAN obtains the lowest RMSE (2.2558×10^{-1}) and rank equal to 1, while the second best model (TCN) presents a considerably higher error (2.7910×10^{-1}). All pairwise comparisons resulted in extremely small p values (ranging from 7.5364×10^{-10} to 1.7764×10^{-15}), and the Holm post hoc correction confirms that all differences remain statistically significant. These results indicate that the RMSE improvements provided by the proposed model are statistically robust.

A consistent behavior can be observed in Table 11, where Opt-EWT-SinWavKAN again achieves the best rank and the lowest MAE (1.4508×10^{-1}). The closest competitor (TCN) shows a substantially higher MAE (2.4864×10^{-1}), corresponding to an increase of approximately 71%. The Wilcoxon test and Holm correction confirm statistical significance against all compared models, demonstrating that the error reductions are consistent across different experimental runs.

Similarly, Table 12 shows that Opt-EWT-SinWavKAN achieves the lowest SMAPE (5.3862×10^{1}), while all other models present errors above 8.1101×10^{1}. The associated p values and Holm corrected values again confirm statistical significance in all comparisons. This confirms that the proposed model provides superior relative error performance in addition to absolute error reductions.

From a methodological perspective, these results suggest that the integration of Empirical Wavelet Transform optimization with the SinWavKAN architecture improves the representation of temporal patterns

Table 11
Wilcoxon signed-rank test against the best mean model for MAE.

Compared model	Avg. Rank	Mean	p-value	Holm p-value	Significant
Opt-EWT-SinWavKAN	1	1.45×10^{-1}	–	–	–
TCN	2	2.49×10^{-1}	7.49×10^{-10}	2.88×10^{-9}	Yes
MLP	3	2.70×10^{-1}	1.54×10^{-12}	2.92×10^{-11}	Yes
TSMixerx	4	2.72×10^{-1}	1.54×10^{-12}	2.92×10^{-11}	Yes
NBEATSx	5	2.79×10^{-1}	1.54×10^{-12}	2.92×10^{-11}	Yes
NBEATS	6	2.79×10^{-1}	1.54×10^{-12}	2.92×10^{-11}	Yes
NHITS	7	2.80×10^{-1}	1.54×10^{-12}	2.92×10^{-11}	Yes
PatchTST	8	2.88×10^{-1}	1.54×10^{-12}	2.92×10^{-11}	Yes
TFT	9	3.00×10^{-1}	1.54×10^{-12}	2.92×10^{-11}	Yes
TimesNet	10	3.08×10^{-1}	7.43×10^{-10}	2.88×10^{-9}	Yes
TiDE	11	3.21×10^{-1}	1.54×10^{-12}	2.92×10^{-11}	Yes
FEDformer	12	3.91×10^{-1}	1.78×10^{-15}	3.73×10^{-14}	Yes
iTransformer	13	4.01×10^{-1}	1.54×10^{-12}	2.92×10^{-11}	Yes
Informer	14	5.19×10^{-1}	1.78×10^{-15}	3.73×10^{-14}	Yes
BiTCN	15	5.25×10^{-1}	7.21×10^{-10}	2.88×10^{-9}	Yes
TSMixer	16	5.49×10^{-1}	1.54×10^{-12}	2.92×10^{-11}	Yes
DilatedRNN	17	6.35×10^{-1}	1.54×10^{-12}	2.92×10^{-11}	Yes
RNN	18	6.37×10^{-1}	1.54×10^{-12}	2.92×10^{-11}	Yes
GRU	19	6.76×10^{-1}	1.54×10^{-12}	2.92×10^{-11}	Yes
LSTM	20	6.97×10^{-1}	1.54×10^{-12}	2.92×10^{-11}	Yes
DeepAR	21	8.01×10^{-1}	1.54×10^{-12}	2.92×10^{-11}	Yes

Table 12
Wilcoxon signed-rank test against the best mean model for SMAPE (%).

Compared model	Avg. Rank	Mean	p-value	Holm p-value	Significant
Opt-EWT-SinWavKAN	1	5.39×10^{1}	–	–	–
TCN	2	8.11×10^{1}	1.78×10^{-15}	3.73×10^{-14}	Yes
TSMixerx	3	8.25×10^{1}	1.54×10^{-12}	2.77×10^{-11}	Yes
NHITS	4	8.27×10^{1}	1.54×10^{-12}	2.77×10^{-11}	Yes
MLP	5	8.39×10^{1}	1.54×10^{-12}	2.77×10^{-11}	Yes
NBEATS	6	8.60×10^{1}	1.54×10^{-12}	2.77×10^{-11}	Yes
NBEATSx	7	8.60×10^{1}	1.54×10^{-12}	2.77×10^{-11}	Yes
FEDformer	8	9.02×10^{1}	1.78×10^{-15}	3.73×10^{-14}	Yes
iTransformer	9	9.43×10^{1}	1.54×10^{-12}	2.77×10^{-11}	Yes
TimesNet	10	9.81×10^{1}	7.55×10^{-10}	2.25×10^{-9}	Yes
TFT	11	9.95×10^{1}	1.54×10^{-12}	2.77×10^{-11}	Yes
Informer	12	9.98×10^{1}	1.78×10^{-15}	3.73×10^{-14}	Yes
BiTCN	13	1.01×10^{2}	7.49×10^{-10}	2.25×10^{-9}	Yes
TSMixer	14	1.02×10^{2}	1.54×10^{-12}	2.77×10^{-11}	Yes
TiDE	15	1.07×10^{2}	1.54×10^{-12}	2.77×10^{-11}	Yes
PatchTST	16	1.08×10^{2}	1.54×10^{-12}	2.77×10^{-11}	Yes
RNN	17	1.09×10^{2}	1.54×10^{-12}	2.77×10^{-11}	Yes
DilatedRNN	18	1.11×10^{2}	1.54×10^{-12}	2.77×10^{-11}	Yes
GRU	19	1.13×10^{2}	1.54×10^{-12}	2.77×10^{-11}	Yes
LSTM	20	1.14×10^{2}	1.54×10^{-12}	2.77×10^{-11}	Yes
DeepAR	21	1.21×10^{2}	1.54×10^{-12}	2.77×10^{-11}	Yes

and reduces forecasting uncertainty. Moreover, the consistency of statistical significance across RMSE, MAE, and SMAPE indicates that the performance gains are not metric dependent but instead reflect a general improvement in predictive capability.

The statistical analysis demonstrates that Opt-EWT-SinWavKAN significantly outperforms all evaluated deep learning baselines, including transformer-based models (Informer, FEDformer, PatchTST, TimesNet, iTransformer), MLP-based architectures (TSMixer, TSMixerx), and recurrent models (LSTM, GRU, RNN, DilatedRNN, DeepAR). The combination of superior ranking, lower mean errors, and statistically significant differences confirms the effectiveness and robustness of the proposed forecasting framework.

5.8. Limitations

While the proposed SinWavKAN model demonstrates superior performance, some limitations should be acknowledged:

- **Computational Complexity:** The combination of EWT denoising, ATPE hyperparameter optimization, and the hybrid SinWavKAN architecture results in higher computational costs during training compared to simpler KAN variants or traditional forecasting models, which may hinder real-time deployment in resource-constrained environments.
- **Interpretability Trade-off:** Although KANs are generally more interpretable than standard deep neural networks, the added complexity of the hybrid sinusoidal-wavelet basis functions may obscure clear insights into individual predictive factors compared to simpler, monotonic models.
- **Limited Exogenous Variables:** The current model focuses on historical volume and inflow data. It does not explicitly incorporate mid- to long-term exogenous variables such as detailed climate model outputs, land-use changes, or long-term climate oscillation indices, which could improve long-horizon forecasts.
- **Generalization to Extreme Events:** While the model handles normal and high-variability periods well, its performance during unprecedented extreme events (e.g., historic floods or mega-droughts) remains untested, as such events are rarely captured in historical training data.

6. Conclusion

This paper presents a novel hybrid Harmonic-Multiscale Kolmogorov–Arnold Network forecasting model, called SinWavKAN, for predicting the useful volume of hydroelectric dams, a critical variable for energy planning and reservoir management. By integrating sinusoidal expansions for capturing global oscillatory patterns and wavelet bases for modeling localized, non-stationary dynamics within the Kolmogorov–Arnold framework, the proposed model achieves an expressive and interpretable architecture tailored for complex hydrological time series.

The proposed methodology incorporates EWT as a preprocessing step to denoise the input signals, effectively isolating informative modes from noise and enhancing the model's ability to learn underlying trends. Furthermore, hyperparameter optimization via the ATPE ensures an optimal configuration of harmonic and wavelet components. Considering the input denoising by EWT and model optimization by APTE, we called our approach Opt-EWT-SinWavKAN.

Experimental results demonstrate the superiority of the proposed model over existing KAN-based models across multiple forecasting horizons and input window sizes. The ablation study confirms the synergistic effect of combining sinusoidal and wavelet representations, which together yield the lowest prediction errors (RMSE, MAE, SMAPE) while maintaining reasonable model complexity. Notably, Opt-EWT-SinWavKAN exhibited significantly slower error growth with increasing forecast horizons, highlighting its robustness for both short- and long-term predictions.

This work contributes an innovative, hybrid KAN architecture that effectively addresses the challenges of non-stationarity, multiscale dynamics, and noise in hydroelectric reservoir volume forecasting. The proposed framework not only advances the state-of-the-art in time series forecasting for energy systems but also offers a flexible, interpretable modeling approach applicable to other domains involving complex temporal dependencies. Future research may explore extending Opt-EWT-SinWavKAN to multivariate forecasting, integrating external climatic variables, and deploying the model in real-time reservoir management systems.

CRediT authorship contribution statement

Stefano Frizzo Stefenon: Writing – original draft, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Laio Oriel Seman:** Writing – original draft. **João Pedro Matos-Carvalho:** Writing – original draft. **Ademir Nied:** Writing – review & editing, Supervision. **Gabriel Villarrubia Gonzalez:** Writing – review & editing. **Kin-Choong Yow:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), funding reference number DDG-2024-00035. Cette recherche a été financée par le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG), numéro de référence DDG-2024-00035. This work is part of the project self-adaptive platform based on intelligent agents for the optimization and management of operational processes in logistic warehouses (PLAUTON), PID2023-151701OB-C21, funded by MCIN/AEI/10.13039/501100011033/FEDER, EU. We thank the Brazilian National Council for Scientific and Technological Development (CNPq) for the grants 305910/2024-8 and 307858/2025-1. This work was also supported by FCT – Fundação para a Ciência e Tecnologia, I.P. under Grants 2023.15441.TENURE.051/CP00003/CT00029 and the LASIGE Research Unit (ref. UID/00408/2025, DOI: https://doi.org/10.54499/UID/00408/2025), and co-financed by FCT and the PRR - Recovery and Resilience Plan by the European Union under the LASIGE Research Unit (UID/PRR/00408/2025, DOI: https://doi.org/10.54499/UID/PRR/00408/2025).

Data availability

The dataset is publicly available at: https://dados.ons.org.br/dataset/ear-diario-por-reservatorio.

References

[1] Operador Nacional do Sistema Elétrico, ONS. Programa mensal da operação (PMO). 2025, https://www.ons.org.br/paginas/energia-no-futuro/programacao-da-operacao. (Accessed: 20 January 2025).

[2] Efstratiadis A, Tsoukalas I, Koutsoyiannis D. Generalized storage-reliability-yield framework for hydroelectric reservoirs. Hydrol Sci J 2021;66(4):580–99. http://dx.doi.org/10.1080/02626667.2021.1886299.

[3] Santos Nd, Lima KC, Spyrides MHC. The dependence of hydropower planning in relation to the influence of climate in northeast Brazil. Plos One 2022;17(1):e0259951. http://dx.doi.org/10.1371/journal.pone.0259951.

[4] Hunt JD, Nascimento A, ten Caten CS, Tomé FMC, Schneider PS, Thomazoni ALR, de Castro NJ, Brandão R, de Freitas MAV, Martini JSC, et al. Energy crisis in Brazil: Impact of hydropower reservoir level on the river flow. Energy 2022;239:121927. http://dx.doi.org/10.1016/j.energy.2021.121927.

[5] Ma X, Hu H, Ren Y. A hybrid deep learning model based on feature capture of water level influencing factors and prediction error correction for water level prediction of cascade hydropower stations under multiple time scales. J Hydrol 2023;617:129044. http://dx.doi.org/10.1016/j.jhydrol.2022.129044.

[6] Ibrahim KSMH, Huang YF, Ahmed AN, Koo CH, El-Shafie A. Forecasting multi-step-ahead reservoir monthly and daily inflow using machine learning models based on different scenarios. Appl Intell 2023;53(9):10893–916. http://dx.doi.org/10.1007/s10489-022-04029-7.

[7] Seman LO, Stefenon SF, Yow K-C, dos Santos Coelho L, Mariani VC. Fourier-enhanced sequence-to-sequence latent graph neural networks for multi-node spatiotemporal forecasting in a hydroelectric reservoir. Eng Appl Artif Intell 2026;167:113939. http://dx.doi.org/10.1016/j.engappai.2026.113939.

[8] Stefenon SF, Seman LO, da Silva LSA, Mariani VC, dos Santos Coelho L. Hypertuned temporal fusion transformer for multi-horizon time series forecasting of dam level in hydroelectric power plants. Int J Electr Power Energy Syst 2024;157:109876. http://dx.doi.org/10.1016/j.ijepes.2024.109876.

[9] Abbas M, Che Y, Khan IU. A novel stacked ensemble framework with the Kolmogorov-arnold network for short-term electric load forecasting. Energy 2025;332:137216. http://dx.doi.org/10.1016/j.energy.2025.137216.

[10] Quanwei T, Guijun X, Wenju X. Cakformer: Transformer model for long-term heat load forecasting based on cauto-correlation and KAN. Energy 2025;324:135460. http://dx.doi.org/10.1016/J.Energy.2025.135460.

[11] Guo S, Wei Z, Yin Y, Zhai X. A physics-guided RNN-kan for multi-step prediction of heat pump operation states. Energy 2025;320:135108. http://dx.doi.org/10.1016/j.energy.2025.135108.

[12] Zhang K, Parolari AJ. Impact of stormwater infiltration on rainfall-derived inflow and infiltration: A physically based surface–subsurface urban hydrologic model. J Hydrol 2022;610:127938. http://dx.doi.org/10.1016/j.jhydrol.2022.127938.

[13] Muniz RN, Stefenon SF, Buratto WG, Nied A, Cardoso R, Yamaguchi CK, Yow K-C. Time series forecasting based on multi-criteria optimization for model and filter selection applied to hydroelectric power plants. Energy 2025;337:138688. http://dx.doi.org/10.1016/j.energy.2025.138688.

[14] Obahoundje S, Diedhiou A, Akpoti K, Kouassi KL, Ofosu EA, Kouame DGM. Predicting climate-driven changes in reservoir inflows and hydropower in côte d'ivoire using machine learning modeling. Energy 2024;302:131849. http://dx.doi.org/10.1016/j.energy.2024.131849.

[15] Shu X, Ding W, Peng Y, Wang Z. Value of long-term inflow forecast for hydropower operation: A case study in a low forecast precision region. Energy 2024;298:131218. http://dx.doi.org/10.1016/j.energy.2024.131218.

[16] Lei K, Chang J, Long R, Wang Y, Zhang H. Cascade hydropower station risk operation under the condition of inflow uncertainty. Energy 2022;244:122666. http://dx.doi.org/10.1016/j.energy.2021.122666.

[17] Zhao Z, Wang Q, Cheng C, Kang Y. A scenario-based multi-agent reinforcement learning approach for efficient solving to long-term optimization of cascade hydropower reservoirs. Energy 2025;340:139356. http://dx.doi.org/10.1016/j.energy.2025.139356.

[18] Zhang J, Wang Y, Wang X, Guo A, Chang J, Niu C, Li Z, Wang L, Ren C. Quantitative analysis and operation strategies for daily-regulation hydropower plants impacted by upstream plant. Energy, Volume=328 2025;136526. http://dx.doi.org/10.1016/j.energy.2025.136526.

[19] Yang P, Jiang Z, Qin H, Zhang Y, Xiao W, Liu G. Application of MS-ICHOA in large scale cascade hydropower stations operation optimization and research on corresponding water storage and release rules. Energy 2025;336:138357. http://dx.doi.org/10.1016/j.energy.2025.138357.

[20] Huang Y, Tan X, Fu J, Deng Z, Liu B. Hydropower vulnerability to drought-flood abrupt alternation under climate change. Energy 2025;340:139212. http://dx.doi.org/10.1016/j.energy.2025.139212.

[21] Zhou S, Wang Y, Su H, Chang J, Huang Q, Li Z. Dynamic quantitative assessment of multiple uncertainty sources in future hydropower generation prediction of cascade reservoirs with hydrological variations. Energy 2024;299:131447. http://dx.doi.org/10.1016/j.energy.2024.131447.

[22] He S, Li B, Li Q, Zheng H, Chen Y. Refining hydropower operation by dynamic control of cascade reservoir water levels with flood season segmentation. Energy 2025;314:134156. http://dx.doi.org/10.1016/j.energy.2024.134156.

[23] Liu Y, Ji C, Wang Y, Zhang Y, Jiang Z, Ma Q, Hou X. Effect of the quality of streamflow forecasts on the operation of cascade hydropower stations using stochastic optimization models. Energy 2023;273:127298. http://dx.doi.org/10.1016/j.energy.2023.127298.

[24] Niu Wj, Luo T, Yao X-r, Gong Jt, Huang Qq, Gao Hy, Feng Zk. Artificial intelligence-based response surface progressive optimality algorithm for operation optimization of multiple hydropower reservoirs. Energy 2024;291:130449. http://dx.doi.org/10.1016/j.energy.2024.130449.

[25] Stefenon SF, Seman LO, da Silva EC, Finardi EC, Coelho LdS, Mariani VC. Hypertuned wavelet convolutional neural network with long short-term memory for time series forecasting in hydroelectric power plants. Energy 2024;313:133918. http://dx.doi.org/10.1016/j.energy.2024.133918.

[26] Granata F, Zhu S, Di Nunno F. Advanced streamflow forecasting for central European rivers: The cutting-edge Kolmogorov-arnold networks compared to transformers. J Hydrol 2024;645:132175. http://dx.doi.org/10.1016/j.jhydrol.2024.132175.

[27] Zhou R, Liu S. A hybrid Kolmogorov-arnold networks-based model with attention for predicting arctic river streamflow. EGUsphere 2025;preprint:1–22. http://dx.doi.org/10.5194/egusphere-2025-3540.

[28] Ren D, Hu Q, Zhang T. EKLT: Kolmogorov-arnold attention-driven LSTM with transformer model for river water level prediction. J Hydrol 2025;649:132430. http://dx.doi.org/10.1016/j.jhydrol.2024.132430.

[29] Bozorgasl Z, Chen Y. Wav-KAN: Wavelet Kolmogorov-arnold networks. 2, 2025, p. 1–19. http://dx.doi.org/10.48550/arXiv.2405.12832, arXiv preprint arXiv:2405.12832.

[30] Larcher JHK, Stefenon SF, dos Santos Coelho L, Mariani VC. Enhanced multi-step streamflow series forecasting using hybrid signal decomposition and optimized reservoir computing models. Expert Syst Appl 2024;255:124856. http://dx.doi.org/10.1016/j.eswa.2024.124856.

[31] Arsenault B, et al. Modified adaptive tree-structured parzen estimator for hyperparameter optimization. 1, 2025, p. 1–21. http://dx.doi.org/10.48550/arXiv.2502.00871, arXiv preprint arXiv:2502.00871.

[32] Jing X, Yang X, Luo J, Zuo G. Exploring Kolmogorov-arnold neural networks for hybrid and transparent hydrological modeling. Environ Model Softw 2025;182:106648. http://dx.doi.org/10.1016/j.envsoft.2025.106648.

[33] Seman LO, Yow KC, Stefenon SF. Sparse mixture of experts enhanced transformer architecture for short-term hydroelectric reservoir volume prediction. Electr Power Syst Res 2026;255:112754. http://dx.doi.org/10.1016/j.epsr.2026.112754.

[34] Ávila L, Silveira R, Campos A, Rogiski N, Freitas C, Aver C, Fan F. Seasonal streamflow forecast in the Tocantins river basin, Brazil: An evaluation of ECMWF-SEAS5 with multiple conceptual hydrological models. Water 2023;15(9):1695. http://dx.doi.org/10.3390/w15091695.

[35] Pillar VD, Overbeck GE. Learning from a climate disaster: The catastrophic floods in southern Brazil. Science 2024;385(6713):eadr8356. http://dx.doi.org/10.1126/science.adr8356.

[36] Stefenon SF, Seman LO, Yamaguchi CK, Coelho LDS, Mariani VC, Matos-Carvalho JP, Leithardt VRQ. Neural hierarchical interpolation time series (NHITS) for reservoir level multi-horizon forecasting in hydroelectric power plants. IEEE Access 2025;13:54853–65. http://dx.doi.org/10.1109/ACCESS.2025.3554446.

[37] Maciel GM, Cabral VA, Marcato ALM, Júnior ICS, Honório LDM. Daily water flow forecasting via coupling between SMAP and deep learning. IEEE Access 2020;8:204660–75. http://dx.doi.org/10.1109/ACCESS.2020.3036487.

[38] Stefenon SF, Ribeiro MHDM, Nied A, Yow KC, Mariani VC, Coelho LS, Seman LO. Time series forecasting using ensemble learning methods for emergency prevention in hydroelectric power plants with dam. Electr Power Syst Res 2022;202:107584. http://dx.doi.org/10.1016/j.epsr.2021.107584.

[39] da Silva EC, Finardi EC, Stefenon SF. Enhancing hydroelectric inflow prediction in the Brazilian power system: A comparative analysis of machine learning models and hyperparameter optimization for decision support. Electr Power Syst Res 2024;230:110275. http://dx.doi.org/10.1016/j.epsr.2024.110275.

[40] Hunt JD, Nascimento A, Zakeri B, Ilyas A, Ramos DS, Kuriqi A, Tolmasquim MT, de Freitas MAV, Brandão R, Wada Y. Optimizing hydropower generation with reservoir level management in humid regions. Energy Rep 2025;13:856–64. http://dx.doi.org/10.1016/J.Egyr.2024.12.064.

[41] Colonetti B, Finardi E, Brito S, Zavala V. Parallel dual dynamic integer programming for large-scale hydrothermal unit-commitment. IEEE Trans Power Syst 2023;38(3):2926–38. http://dx.doi.org/10.1109/TPWRS.2022.3187059.

[42] Leng Z, Chen L, Yi B, Liu F, Xie T, Mei Z. Short-term wind speed forecasting based on a novel kaninformer model and improved dual decomposition. Energy 2025;322:135551. http://dx.doi.org/10.1016/j.energy.2025.135551.

[43] Gao R, Du L, Duru O, Yuen KF. Time series forecasting based on echo state network and empirical wavelet transformation. Appl Soft Comput 2021;102:107111. http://dx.doi.org/10.1016/j.asoc.2021.107111.

[44] Gilles J. Empirical wavelet transform. IEEE Trans Signal Process 2013;61(16):3999–4010. http://dx.doi.org/10.1109/TSP.2013.2265222.

[45] Zhang Z, Cheng Y, Liu NC. Comparison of the effect of mean-based method and z-score for field normalization of citations at the level of web of science subject categories. Scientometrics 2014;101:1679–93. http://dx.doi.org/10.1007/s11192-014-1294-7.

[46] Shi Z, Zhang G, Lu C, Zhou X, Lv X. Dynamic spatio-temporal graph-enhanced kanformer for high-fidelity ultra-short-term wind power forecasting. Energy 2025;337:138590. http://dx.doi.org/10.1016/j.energy.2025.138590.

[47] Zhao Q, Wang S, Chen Y, Liu J, Sun Y, Su T, Li N, Fang J. A hybrid framework for short-term load forecasting based on optimized InMetra boost and bilstm. Energy 2025;328:136582. http://dx.doi.org/10.1016/j.energy.2025.136582.

[48] Mahdi WA, Alhowyan A, Obaidullah AJ. Advanced hybrid intelligence evaluation of nanomedicine delivery to various organs using machine learning and adaptive tree structured parzen estimator. Sci Rep 2025;15(1):28188. http://dx.doi.org/10.1038/s41598-025-13028-w.

[49] Stefenon SF, Yow K-C. Adaptive filter-driven optimized attention-based CNN-LSTM for load forecasting in microgrids. Results Eng 2025;28:107470. http://dx.doi.org/10.1016/j.rineng.2025.107470.

[50] Somvanshi S, Javed SA, Islam MM, Pandit D, Das S. A survey on kolmogorov-arnold network. ACM Comput Surv 2025;58(2):1–35. http://dx.doi.org/10.1145/3743128.

[51] Essahraui S, Lamaakal I, Maleh Y, El Makkaoui K, Bouami MF, Ouahbi I, Elmannai H, Abd El-Latif AA. FastKAN-DDD: A novel fast Kolmogorov-arnold network-based approach for driver drowsiness detection optimized for tinyml deployment. PLos One 2025;20(11):e0332577. http://dx.doi.org/10.1371/journal.pone.0332577.

[52] Miraç KÖKÇAM Ö, Ucar F. Bit-HyMLPKANClassifier: A hybrid deep learning framework for human peripheral blood cell classification using big transfer models and Kolmogorov–Arnold networks. Adv Intell Syst 2026;8(1):2500387. http://dx.doi.org/10.1002/aisy.202500387.

[53] Abdallah Mahmoud A, Pester A, Muttardi MM, Andres F, Tanabe S, Greneche N, Ali HH. Cheby-KANs: Advanced Kolmogorov–Arnold networks for applying geometric deep learning in quantum chemistry applications. IEEE Access 2025;13:130525–34. http://dx.doi.org/10.1109/ACCESS.2025.3566551.

[54] Zhang J, Dou Y, Zhang P, Shu X, Lei M. Radial basis function Kolmogorov–arnold network for coal calorific value prediction using portable near-infrared spectroscopy. Processes 2025;13(11):3623. http://dx.doi.org/10.3390/pr13113623.

[55] Xue Z, Ye Z, Wang B, Zhang Y, Zheng C, Peng J. Enhancing hyperspectral image classification with spectral–spatial convolutional Fourierkan transformer. IEEE J Sel Top Appl Earth Obs Remote Sens 2026;19:3673–86. http://dx.doi.org/10.1109/JSTARS.2025.3649558.

[56] Lu Y, Zhang R, Suo Z, Zhang J, Liu L, Min G. CFTD: Core fusion time series dense encoder for intelligent prediction with edge AI in social IoT systems. IEEE Internet Things J 2025;12(17):34979–91. http://dx.doi.org/10.1109/JIOT.2025.3569950.

[57] Muniz RN, Buratto WG, Gonzalez GV, Seman LO, Costa VJ, Nied A. Optimized hybrid neural hierarchical interpolation time series with stl for flow forecasting in hydroelectric power plants. Sci Rep 2026;16:5002. http://dx.doi.org/10.1038/S41598-025-34847-X.

[58] Carvalho JPM, Stefenon SF, Yow KC, Ovejero RG, Leithardt VRQ. N-BEATS neural network applied for insulator fault prediction considering EMD methods. In: International conference on electrical, computer, communications and mechatronics engineering, vol. 5, Zanzibar, Tanzania: IEEE; 2025, p. 1–6. http://dx.doi.org/10.1109/ICECCME64568.2025.11277686.

[59] Huang R, Wang C, Lin H, Zhang H, Ma H. CWGWO-n-BEATSx: An improved time-series prediction method with multiple external variables for situation prediction. IEEE Internet Things J 2025;12(18):37281–300. http://dx.doi.org/10.1109/JIOT.2025.3581958.

[60] Mahaseni B, Faizal ERM, Raj RG. Spotting football events using two-stream convolutional neural network and dilated recurrent neural network. IEEE Access 2021;9:61929–42. http://dx.doi.org/10.1109/ACCESS.2021.3074831.

[61] Chen B, Zhu L, Hu L, Zhang R, Wu Y, Li H, Wen X, Zhang Y, Gao K. Adaptive energy management of electric vehicles via attention-enhanced LSTM networks for load power demand prediction. Energy 2025;344:139797. http://dx.doi.org/10.1016/J.Energy.2025.139797.

[62] Wang J, Yang Q, He L, Liu X, Fu L, Zhang C. Dynamic scenario simulations of carbon emission peaking and mitigation pressure with CNN-gru-attention model in China's construction sector. Energy 2026;344:140132. http://dx.doi.org/10.1016/J.Energy.2026.140132.

[63] Deng J, Xiao Z, Zhao Q, Zhan J, Tao J, Liu M, Song D. Wind turbine short-term power forecasting method based on hybrid probabilistic neural network. Energy 2024;313:134042. http://dx.doi.org/10.1016/j.energy.2024.134042.

[64] Xu R, Fang H, Zeng H, Wu B. A novel interpretable wind speed forecasting based on the multivariate variational mode decomposition and temporal fusion transformer. Energy 2025;331:136497. http://dx.doi.org/10.1016/j.energy.2025.136497.

[65] Zhu W, Li C, Xu Y, Yang W, Xie C. High accuracy and adaptability of PEMFC degradation interval prediction with informer-GPR under dynamic conditions. Energy 2024;307:132781. http://dx.doi.org/10.1016/j.energy.2024.132781.

[66] Pan M, Cao X, Fu C, Liao S, Zhou X, Guan W. Emission prediction and optimization of methanol/diesel dual-fuel engines based on itransformer-bigru and NSGA-III. Energy AI 2025;19:100466. http://dx.doi.org/10.1016/j.egyai.2024.100466.

[67] Hou G, Zhou Y, Zhang F. Data-driven modeling of ultra-supercritical unit based on improved patchtst network with multi-task balancing. Energy 2025;330:136958. http://dx.doi.org/10.1016/j.energy.2025.136958.

[68] Wen Y, Pan S, Li X, Li Z, Wen W. Improving multi-site photovoltaic forecasting with relevance amplification: DeepFEDformer-based approach. Energy 2024;299:131479. http://dx.doi.org/10.1016/J.Energy.2024.131479.

[69] Zhan Y, Wang X, Xu Y, Li W. A hybrid TCN-LSTM-attention framework for multi-scenario short-term photovoltaic power forecasting incorporating physics-informed neural network strategy. Energy 2026;344:139968. http://dx.doi.org/10.1016/j.energy.2026.139968.

[70] Ding L, Li Q, Chen X. Short-term photovoltaic power prediction based on fcformer-bitcn method. IEEE Trans Instrum Meas 2025;74:1–13. http://dx.doi.org/10.1109/TIM.2025.3638927.

[71] Gong J, Qu Z, Zhu Z, Xu H. Parallel TimesNet-bilstm model for ultra-short-term photovoltaic power forecasting using STL decomposition and auto-tuning. Energy 2025;320:135286. http://dx.doi.org/10.1016/j.energy.2025.135286.