Analysis of Ensembles Applied to Time Series Forecasting of Hydroelectric Dams

Jheymesson Apolinário Cavalcanti
J.A. Cavalcanti*
jheymesson.cavalcanti@unicap.br
jac@ecomp.poli.br
Universidade de Pernambuco
Recife, Pernambuco, Brazil

Roberta Andrade de Araujo Fagundes
R.A.A. Fagundes
Universidade de Pernambuco
Recife, Pernambuco, Brazil
roberta.fagundes@upe.br

Abstract

Context: Dams are essential for electricity generation in Brazil. Given the countrys vast territorial extent, it is crucial to optimize the use of its energy sources, particularly the hydrological potential of rivers. Problem: Flow forecasting in river basins is critical for the safety and efficiency of operations. Underestimation errors can increase the risk of dam overflow and failure, while overestimations compromise water storage, affecting supply and power generation during dry periods. Solution: This study proposes a dynamic selection method of machine learning models to achieve more accurate and stable flow forecasts. IS Theory: The research is based on Dynamic Capabilities Theory, emphasizing the ability for internal reconfiguration to address complex issues such as time series forecasting. Method: Data were initially normalized and temporally adjusted to ensure consistency. A diverse set of base models was selected based on their intrinsic characteristics, then integrated into a dynamic ensemble framework. Performance was evaluated using MAPE and RMSE metrics, allowing comparative analysis between the proposed dynamic ensemble and traditional deep learning models. Results: The dynamic selection approach showed significant improvements, achieving errors of 23.88 (MAPE) and 852.97 (RMSE), outperforming traditional and deep learning models. Contributions and Impact in the IS Field: This work demonstrates that dynamic model selection is a superior and promising approach compared to isolated model use, offering a significant advancement for flow forecasting in dam systems and contributing to more efficient water resource management in the Brazilian context.

CCS Concepts

• **Computing methodologies** → **Ensemble methods**; • **Information systems** → *Decision support systems.*

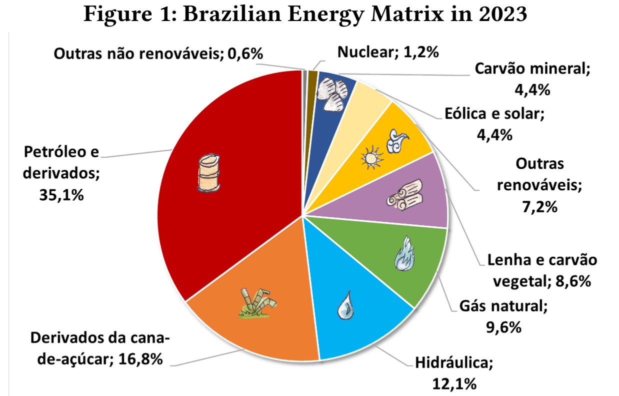
Keywords

Time series forecast, Ensembles, Dynamic selection

1 Introduction

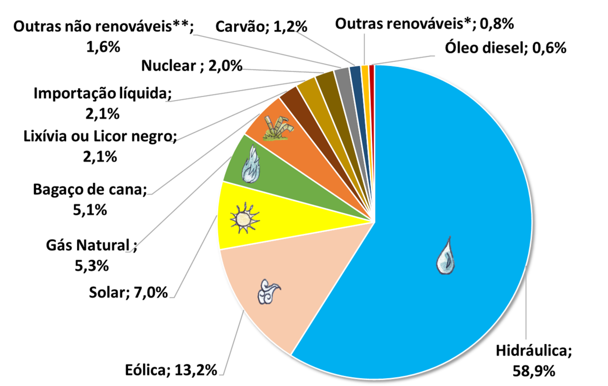
Hydroelectric dams are essential for energy production in Brazil. Their significance is underscored by the fact that hydropower resources accounted for 12.1% of the country's energy resources in 2023, as shown in Figure 1, which presents Brazil's energy matrix as a reference for available resources for electricity generation. Additionally, the demand and use of these energy resources represented 58.9% of the Brazilian electrical matrix in 2023, as indicated in Figure 2. This importance is attributed to several factors, including the renewability of hydropower resources and Brazil's established infrastructure, both in terms of water distribution through its extensive river network and the numerous hydroelectric facilities already constructed.

Figure 1: Brazilian Energy Matrix in 2023



Source: [4]

Figure 2: Brazilian Electrical Matrix in 2023



Source: [4]

Hydroelectric plants play a crucial and responsible role in Brazilian society, generating over half of the country's energy supply and managing extensive resources. Mismanagement of these resources can lead to serious issues. Underestimated inflows can prevent adequate water release, increasing dam levels and the risk of overflow or, in extreme cases, dam failure. Such events could result in downstream flooding, impacting the safety of local communities, infrastructure, and river ecosystems [20]. Additionally, unplanned floods can cause significant economic losses, affecting residences, industries, transportation networks, and agriculture. Therefore, accurate flow forecasting is essential and any improvement in prediction precision is of utmost importance for dam management [20].

1.1 Problem Statement

Time series, in a general context, are complex and unique, making them challenging to forecast. Often, models that perform well in training, validation, and testing stages may still yield poor results in real-world scenarios. This can happen for various reasons, such as changes in the nature of the series due to external factors, the predictive model entering an undertrained region, or the model becoming outdated over time, necessitating re-modeling [10].

In general, relying on a single predictive model is not an ideal solution, as it may adhere to a narrow view of the data. As the data evolves, a single model may fail to adapt to these changes [21]. Therefore, ensemble methods have gained prominence, especially those employing model selection strategies. Static ensembles maintain a fixed combination of models throughout the forecasting process, while dynamic ensembles adapt their composition or weighting based on the performance of models over time. This work will compare the performance of static and dynamic ensemble strategies and contrast their results with those obtained from individual base models traditionally used in time series forecasting. The goal is to evaluate whether the flexibility provided by dynamic selection can yield significant improvements over static selection and single models, particularly in non-stationary environments.

1.2 Justification

Traditional models tend to experience performance degradation over time. To address this, various application fields have adopted the use of multiple combined models, known as ensembles, rather than relying on a single model. Ensembles are the combination of multiple base models [1].

Ensembles save significant effort in the search and modeling stages, as it is common practice to exhaustively test multiple models until an optimal model is situationally identified. However, the most frustrating aspect is that, despite this extensive effort, a single model alone does not ensure long-term performance stability.

This is why ensembles that dynamically select base models are particularly advantageous for this field of application. When a base model's performance begins to degrade, it is replaced by a more promising model for that temporal context [1].

2 Literature Review

2.1 Time Series Forecasting

Time series forecasting is a critical field across multiple disciplines, aiming to estimate future values based on observed data over time. Its application is crucial in scenarios where anticipating changes can guide strategic and operational decisions, such as in economics, resource planning, risk management, and demand forecasting [8]. However, time series forecasting is a complex challenge due to the dynamic and multifaceted nature of the variables involved, which often follow nonlinear patterns and exhibit interdependent relationships that are difficult to model [5].

One of the main difficulties in time series forecasting is the presence of trend, seasonality, and cyclical components that interact in a non-trivial way and often vary over time [13]. These variations can be influenced by external factors, such as economic, political, or environmental changes, introducing breaks or alterations in the series patterns. Additionally, noise in the data, that is, random variation without any identifiable pattern, can obscure underlying relationships and increase forecast uncertainty [6].

Another obstacle is the intrinsic temporal dependence of observations, where past events directly influence future behavior. Effectively modeling these relationships requires the capability to capture both short- and long-term dependencies and to identify which factors hold predictive relevance in each period. Furthermore, the robustness of predictive models is constantly challenged by structural shifts in data, which can invalidate previously effective methods, requiring an adaptive approach to accommodate new patterns. Thus, time series forecasting demands advanced and flexible methodologies that allow for a deep understanding of the data and the ability to adapt to their temporal complexity and variability [15].

Another crucial point of this proposal is not to combine the final models but to select a single final model to represent the ensemble as a whole. This approach is applied to the dynamic selection method by recent error window, making the model simpler and more efficient.

2.2 Base Models

The ARIMA (AutoRegressive Integrated Moving Average) model is widely used for modeling and forecasting time series, particularly effective for data exhibiting seasonal patterns and trends over time [24]. It combines three main components: the AutoRegressive (AR) term, representing the relationship between the current value and past values of the series; the Integrated (I) term, referring to differencing, a process that makes the series stationary by calculating the difference between consecutive values; and the Moving Average (MA) term, which uses the average of past errors to improve predictions. The ARIMA model equation is given by (1), where Y_t is the current value, c is the model constant, $\phi_1, \ldots, \phi_p$ are the autoregressive components, $\theta_1, \ldots, \theta_q$ are the moving average coefficients, and e_t is the random error.

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} + e_t \tag{1}$$

The Multilayer Perceptron (MLP) is a type of artificial neural network primarily used for classification and regression problems. It consists of multiple layers of neurons: an input layer, one or more hidden layers, and an output layer, where each layer is fully connected to the next. MLP uses a nonlinear activation function, such as ReLU (Rectified Linear Unit) or sigmoid, to allow the network to learn complex, nonlinear patterns. The learning process in MLP is achieved through backpropagation, which adjusts the weights ω_{ij} between neurons to minimize the error between the network's predictions and the actual values, with x_i representing inputs and b_j as the bias. The output of each neuron y_j in a hidden or output layer is described by equation (2), where f is the activation function of the artificial neuron [25].

$$y_j = f\left(\sum_{i=1}^{n} w_{ij}x_i + b_j\right) \tag{2}$$

The Support Vector Regression (SVR) is primarily used for regression problems, being an adaptation of the Support Vector Machine (SVM) for forecasting continuous values. SVR aims to find a function that fits the data such that the prediction error is within a margin ϵ, while minimizing model complexity. The SVR prediction function is represented by (3), where x_i represents support vectors, α_i and α_i^* are Lagrange multipliers adjusted during training, $K(x_i, x)$ is the kernel function that calculates the similarity between samples x_i and x, and b is the bias term. The kernel function enables SVR to capture nonlinear relationships in the data, commonly using linear, polynomial, or RBF (Radial Basis Function) kernels [18].

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(x_i, x) + b \tag{3}$$

The Extreme Learning Machine (ELM) is a single-layer neural network, mainly used for classification and regression tasks, known for its computational efficiency. In ELM, the weights ω_i between the input layer and the hidden neurons, as well as the biases b_i, are randomly set and remain fixed, while only the weights β_i of the output layer are adjusted to minimize the error. The ELM prediction function is represented by (4), where $f(x)$ is the predicted output for an input x, g is the activation function (such as sigmoid, ReLU, or hyperbolic tangent), and N is the number of hidden layer neurons. The simplicity of ELM training, where an analytical solution determines the output weights, makes it extremely fast and efficient, especially on large datasets [16].

$$f(x) = \sum_{i=1}^{N} \beta_i g(w_i \cdot x + b_i) \tag{4}$$

Finally, the Long Short-Term Memory (LSTM) is a recurrent neural network architecture designed to model long-term dependencies in sequential data, widely used in time series forecasting and natural language processing. LSTM overcomes gradient vanishing issues common in recurrent networks by introducing a "memory cell" with three gates controlling the flow of information: input, forget, and output gates [9]. The memory cell update is represented by equations (5):

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{5}$$

Where x_t is the current input, h_{t-1} is the previous hidden state, C_t is the cell state, f_t, i_t, and o_t are the forget, input, and output gate activations, W and b are weights and biases, σ is the sigmoid function, and tanh is the hyperbolic tangent. The LSTM effectively learns which information to retain and discard over sequences, making it powerful for capturing long-term patterns in sequential data.

2.3 Ensembles

The production of an ensemble for time series forecasting involves three main phases: generation, selection, and composition. Each of these stages plays a critical role in developing a robust and accurate model capable of capturing the inherent complexity and variability of temporal data [7].

(1) Generation: The generation phase is responsible for creating a diverse set of forecasting models that will serve as the base for the ensemble. In the context of time series, this diversity is essential to capture different patterns, such as trends, seasonality, and cyclic fluctuations, which may not be adequately modeled by a single model. Generation can be carried out in several ways, including using different learning algorithms (e.g., neural network-based methods, statistical methods, regression models), varying hyperparameters for the same algorithm, or even using different data approaches, such as variations in time windows. The diversity of generated models enhances the ensemble's ability to handle distinct patterns and reduces the risk of overfitting to specific patterns.
(2) Selection: The selection phase is crucial for identifying the most effective models from the generated set, adjusting the ensemble to maximize its predictive accuracy. In time series forecasting, this selection can be performed in a static or dynamic manner. Static selection considers the performance of models on a fixed validation dataset, selecting those with the best performance for the final ensemble. Dynamic selection, on the other hand, adapts the selected model set based on recent data behavior, such as the recent error window or K-Nearest Neighbors (KNN) approach, which prioritizes models that have proven more effective in capturing patterns similar to the current period. Proper model selection enables finer adaptation of the ensemble to the dynamic nature of time series.
(3) Composition: The composition phase is the process of combining the predictions of the selected models to generate a final ensemble forecast. There are several approaches to combining models, with the most common being simple averaging, weighted averaging, and more complex combinations, such as stacked regression or adaptive weighting networks. Weighted averaging is particularly useful in time series, as it allows greater influence to be attributed to models that have proven more accurate in recent forecasts. In more advanced contexts, combination techniques based on adaptive learning allow the ensemble to adjust over time, assigning dynamic weights based on performance at each point in the series. The composition phase is essential for balancing each model's contribution so that the ensemble provides not only high accuracy but also robustness and stability throughout the series.

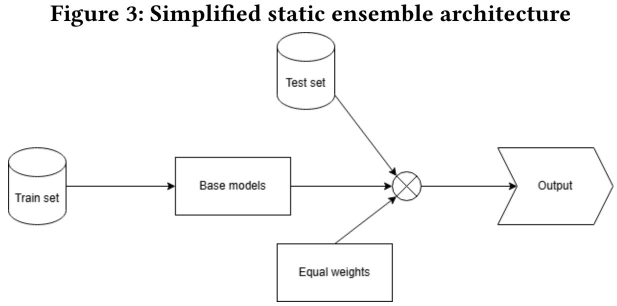
In summary, the three phases of producing an ensemble for time series forecasting—generation, selection, and composition—work in synergy to build a predictive model capable of handling the complexity of temporal data. Diversity in generation, adaptability in selection, and robustness in composition are key elements to achieving accurate and reliable predictions, especially in dynamic and challenging data environments. This work focuses on the selection phase of base models, specifically on pruning models for optimized ensemble performance.

2.3.1 *Static Selection.*

In time series forecasting, the use of heterogeneous ensembles composed of different predictive models is an effective strategy to improve forecast accuracy and robustness. The combination of models with complementary characteristics allows for the capture of various patterns and dynamics within the time series, such as trends, seasonality, and random fluctuations, which may be better represented by specific models. In the context of heterogeneous ensembles, two widely used approaches for selecting models for the ensemble are fixed static selection and weighted static selection [2].

Fixed static selection involves pre-defining the models that will compose the ensemble without dynamic adjustments over time follows a simplified diagram of its architecture in the Figure(3). The models are selected based on their historical performance on the training dataset, and this composition remains unchanged throughout the forecasting process. This approach is advantageous in terms of simplicity and computational efficiency, as it avoids frequent recalibrations and allows for a more straightforward ensemble implementation. However, a challenge with this approach is that it may be less adaptive to changes in the characteristics of the time series, such as structural shifts or pattern changes over time, which may necessitate adjustments in the composition of selected models [11].

Figure 3: Simplified static ensemble architecture

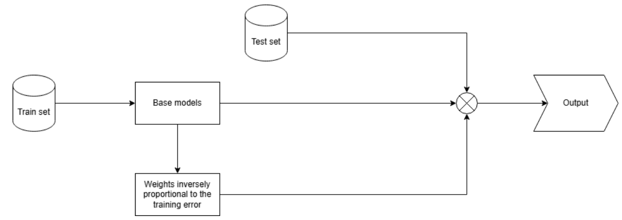


Source: Author

On the other hand, weighted static selection advances by assigning fixed weights to each model in the ensemble, also based on their historical performance. This weighting aims to reflect each model's relative contribution to the final forecast, with better-performing models receiving higher weights as shown in the Figure(4). The final forecast is then calculated as a weighted average of the individual model forecasts. This approach can capture the time series behavior more robustly by adjusting each model's influence based on its relative performance, allowing for greater flexibility without the need for dynamic weight changes during the forecasting process [22].

While both static approaches do not respond to sudden data variations, they offer practical implementation advantages and lower computational costs, making them viable options for scenarios where the stability of time series patterns is reasonably assured. In summary, fixed and weighted static selection provide a solid methodological basis for constructing efficient heterogeneous ensembles in time series forecasting, as long as data stability conditions are observed and the limitations of adaptation to new dynamics are considered [22].

Figure 4: Simplified architecture of the weighted static ensemble



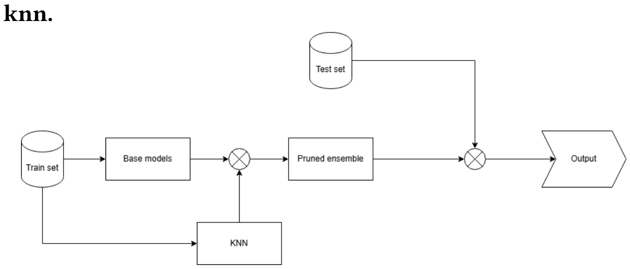
Source: Author

2.3.2 *Dynamic Selection.*

Dynamic selection of a heterogeneous ensemble in time series forecasting is an advanced approach aimed at improving forecast accuracy and adaptability. Instead of using a fixed combination of models throughout the entire series, as in static ensembles, dynamic selection adjusts each model's contribution as the data behavior changes. This approach allows for responses to temporal changes, such as seasonal shifts, structural breaks, and short-term patterns, providing more robust and accurate forecasts in complex and dynamic scenarios.

A popular technique for performing this dynamic selection is the K-Nearest Neighbors (KNN), which selects, at each forecast instance, the models in the ensemble that performed best on similar past observations. In the context of time series forecasting, KNN can be applied by identifying historical periods in which the series conditions (such as trend and seasonality) are close to the current conditions. Models that achieved the lowest errors in these similar periods are then selected to compose the ensemble. This approach is powerful in scenarios where the series exhibits recurring patterns, allowing the ensemble to automatically adapt to historical cycles and capture seasonal dynamics as shown in the Figure(6).

Figure 5: Simplified architecture of dynamic ensemble by knn.

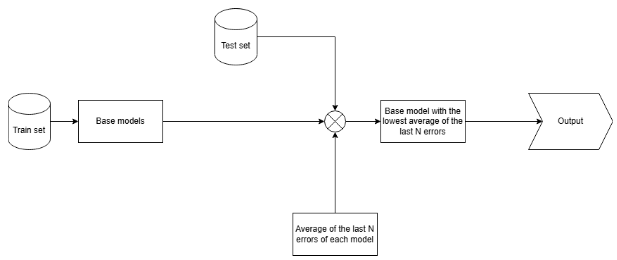


Source: Author

Another dynamic selection technique is the recent error window [17], which considers the recent performance of models to determine which ones should be selected. In this method, each model is evaluated based on the errors it made in an immediately preceding observation window. Models with the lowest average error in this recent window are chosen to compose the ensemble, as they have demonstrated better adaptation to the most recent series behavior. This method is particularly effective in time series that frequently change patterns, allowing the ensemble to quickly adapt to new conditions by prioritizing models that respond well to the most current data behavior.

This method strongly aligns with the IS theory of dynamic capabilities, which emphasizes adaptability and reconfiguration in response to rapidly changing environments. By continuously selecting the model that performs best in the recent error window, the proposed architecture applies this theory by making a single-model selection at each prediction interval as shown in the Figure(??). Specifically, it identifies and utilizes only the model with the lowest average error in the recent window as the sole representative of the ensemble. This single-model selection approach simplifies the ensemble output, reduces computational overhead, and ensures that the forecasting process remains both efficient and highly adaptive to evolving data patterns, embodying the core principles of dynamic capabilities.

Figure 6: Simplified architecture of dynamic ensemble by knn.



Source: Author

2.4 Contributions

Forecasting models frequently face the challenge of abrupt or gradual changes in the nature of data, without prior indication, either due to exogenous interferences or the stochastic dynamics of the underlying processes. These changes, referred to as concept drifts or shifts, are often identified late, compromising the accuracy of previously trained models. As a result, such models may become obsolete and require reconfiguration or replacement to maintain predictive efficiency.

In this context, heterogeneous ensembles have emerged as a robust solution, as in scenarios of data instability, the pruning and updating policy of the ensemble itself can automatically adapt the composition of the model set, mitigating the impacts of concept changes.

The literature shows that the combination of diverse models tends to outperform individual models, provided it is supported by an effective strategy for selecting and removing suboptimal components. Moreover, in highly complex problems, merely eliminating low-performing predictors already constitutes a tangible benefit for the system as a whole.

Another widely discussed aspect in the literature is that dynamic model selection approaches usually outperform static strategies in terms of accuracy and robustness [12]. The main contribution of this work lies in the proposal of an adaptive method that monitors the region of the vector space where forecasts are being generated, identifying situations in which the performance of the current model becomes unsatisfactory. Upon detecting these unfavorable regions, the need for model replacement is signaled; however, this replacement is carried out gradually, allowing the system to transition smoothly between models, ensuring the stability and continuity of forecasts.

This gradual transition process aims to prevent abrupt performance oscillations in the system, contributing to a more robust approach adaptable to changes in data patterns. This strategy aligns with recent trends in the literature, which emphasize the importance of adaptive mechanisms and smooth transitions in the dynamic selection of predictors in non-stationary data environments.

2.5 Stages

This section outlines the research methodology employed in this study, focusing on estimation using artificial neural networks (ANN). Fig. 7 illustrates the four key phases in addressing the problem: problem understanding (Phase 1), data processing (Phase 2), problem modeling (Phase 3), and finally, evaluation and discussion of results (Phase 4).

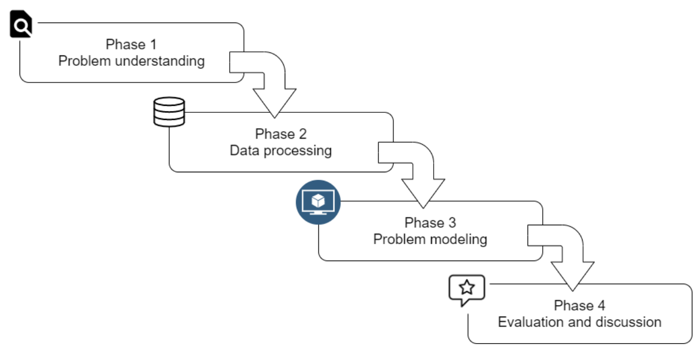


Figure 7: Division of methodology into four research phases.
Source: Author

2.6 Problem Understanding

As previously mentioned, the context of time series forecasting is highly complex and subject to rapid changes over short periods. Applying the Dynamic Capabilities Theory allows the ensemble model to remain stable and robust, with strong generalization potential through model selection within the ensemble, a method that tends to reduce error metrics and presents an excellent long-term solution [23].

2.7 Data Processing

The data utilized consists of monthly average flow rates from three Brazilian hydroelectric dams: Sobradinho, Tucuruí, and Três Marias, covering historical records from January 1931 to February 2017.

Given the characteristics of Brazilian hydrographic basins, their flow rates naturally exhibit frequent fluctuations between high and low values. This variability is largely attributed to seasonal climatic patterns, particularly alternating periods of floods and droughts. Consequently, it is an inherent feature of these datasets to present both extreme values and a high standard deviation, as indicated in

Table 1. This table provides a statistical summary of the time series, detailing the dataset name, minimum and maximum observed values, mean flow rate, and standard deviation (SD), respectively.

Initially, the data were normalized to a scale between 0 and 1 using linear normalization. This step prevents data-sensitive models from prioritizing higher values. Subsequently, a 14-step lag was applied to predict the next step, constituting a sliding window of 14 steps for one-step-ahead forecasting, selected through exhaustive testing to provide sufficient information for model learning.

The data were then divided into 75% for training and 25% for testing, ensuring temporal order preservation to respect the sequential nature of time series data.

Table 1: Statistical summary of the datasets.

Dataset	Minimum	Maximum	Mean	SD
Sobradinho	226.62	15676	2606.23	1941.17
Tucuruí	1269	51539	10935.24	9182.30
Três Marias	40.48	4435	677.31	600.51

2.8 Modeling

2.8.1 Base Models. The base models, or monolithic models, were selected based on their unique characteristics as follows:

- **ARIMA**: Extensively used in the literature, it performs well on series with linear dependencies.
- **Multi-Layer Perceptron (MLP)**: A stochastic-based model that is effective with non-linear dependencies.
- **Support Vector Regression (SVR)**: An adaptation of Support Vector Machine (SVM) for regression tasks; unlike MLP, it is non-stochastic and handles non-linear data well.
- **Extreme Learning Machine (ELM)**: Known for high performance and excellent generalization, particularly effective with non-linear dependencies.
- **Long Short-Term Memory (LSTM)**: A deep learning model with gated memory units, excelling in capturing long-term dependencies in sequential data.

2.8.2 Ensembles. There are two possible approaches for choosing base models to construct an ensemble: homogeneous and heterogeneous. The homogeneous ensemble uses variations of a single model, while the heterogeneous ensemble comprises different base models, each contributing uniquely to the final result. For this study, a heterogeneous ensemble was chosen due to its superior generalization capability.

The heterogeneous ensemble employed includes the base models described in Section 2.8.1. The choice of a heterogeneous ensemble ensures that diverse learning biases contribute to the final result, leading to a more consistent and fair outcome.

2.9 Evaluation Metrics

This study employs two widely-used performance metrics in time series forecasting literature: the Mean Absolute Percentage Error (MAPE) and the Root Mean Square Error (RMSE), as defined in Equations (6) and (7), respectively.

$$\text{MAPE} = \frac{100}{n} \sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \tag{6}$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (y_t - \hat{y}_t)^2} \tag{7}$$

Lower values indicate better model performance, with MAPE providing error as a percentage of expected output and RMSE measuring the squared error difference between predicted and actual outputs. These metrics were chosen because they are among the most widely used in the literature, ensuring consistency and comparability with previous studies.

2.10 Optimization

Lag selection was performed using exhaustive testing, yielding a 14-step window for one-step-ahead forecasting. For hyperparameter optimization of machine learning-based models, Particle Swarm Optimization (PSO) was utilized. PSO is a swarm intelligence-based optimization technique, widely used to tune hyperparameters of complex models such as LSTM, SVR, MLP, and ELM in time series forecasting [14]. PSO models a group of particles exploring the solution space to find optimal configurations. Each particle represents a potential solution, iteratively updating its position based on individual and group experiences, balancing exploration and exploitation [19].

For each model's hyperparameters, PSO updates particle positions according to Equation (8), where $v_i(t)$ is the velocity of particle i at time t, $x_i(t)$ is the current position, p_i represents the particle's best-known position, g denotes the global best position, w is the inertia factor, and c_1, c_2 are learning coefficients with random factors r_1 and r_2. This update enables particles to adjust their positions based on the best local and global solutions, converging to an optimized hyperparameter configuration.

$$v_i(t+1) = w \cdot v_i(t) + c_1 \cdot r_1 \cdot (p_i - x_i(t)) + c_2 \cdot r_2 \cdot (g - x_i(t)) \tag{8}$$

Applying PSO to models such as LSTM, SVR, MLP, and ELM, each particle's configuration is evaluated using error metrics like RMSE and MAPE. For LSTM and MLP, PSO adjusts parameters such as the number of layers and learning rate; for SVR, it optimizes the regularization parameter and margin of error; for ELM, it determines the ideal number of neurons and activation function. At the end, PSO identifies the configuration that maximizes predictive accuracy and generalization capability, enhancing forecast robustness for time series data.

In contrast, the ARIMA model was optimized using the `autoarima` function, which selects and calculates optimal hyperparameters for the dataset [3].

3 Results

The proposed modification to the dynamic selection model using a recent error window approach demonstrated highly promising results, effectively identifying the optimal models at critical moments and thereby reducing prediction errors. It is noteworthy that the datasets in this study have high magnitudes, where even slight percentage errors translate to substantial discrepancies in practical applications, underscoring the impact of achieving low MAPE and RMSE values.

Applying the models across three distinct datasets yielded consistent outcomes. For the Sobradinho dataset, the dynamic ensemble by recent error window achieved 24.79 for MAPE and 825.82 for RMSE, as detailed in Table 2. Similarly, for the Três Marias dataset, the approach achieved 36.72 and 358.07 for MAPE and RMSE, respectively, as shown in Table 3. Finally, for the Tucuruí dataset, the ensemble reached MAPE and RMSE values of 18.81 and 3155.05, respectively, as indicated in Table 4. In all three cases, the modified dynamic selection based on the recent error window outperformed other models and ensembles, especially in RMSE, where the reductions were substantial. Although MAPE improvements appeared modest, their impact is significant given the magnitude of the data, and the pronounced RMSE gains reinforce the robustness of the approach.

Model	MAPE	RMSE
ARIMA	61.42	1084.37
MLP	31.51	877.39
SVR	26.20	860.01
ELM	46.79	1173.05
LSTM	37.25	981.10
Static Ensemble	42.68	936.04
Weighted Static Ensemble	40.78	921.48
Dynamic Ensemble by KNN	34.38	927.79
Dynamic Ensemble by Window	**24.79**	**825.82**

Table 2: Performance Comparison of Different Models (MAPE and RMSE) on the Sobradinho Dataset

Model	MAPE	RMSE
ARIMA	94.67	425.03
MLP	82.83	422.98
SVR	38.68	370.16
ELM	51.92	371.68
LSTM	77.30	394.47
Static Ensemble	70.29	387.90
Weighted Static Ensemble	65.32	381.85
Dynamic Ensemble by KNN	46.24	371.42
Dynamic Ensemble by Window	**36.72**	**358.07**

Table 3: Performance Comparison of Different Models (MAPE and RMSE) on the Três Marias Dataset

The proposed dynamic selection approach consistently outperformed the other models across all three datasets. It is noteworthy that although not all ensemble methods showed substantial improvements, each maintained a significant degree of consistency, highlighting the value of ensemble stability in temporal prediction.

The forecast results of the proposed ensemble for each dataset are illustrated in Figures 8, 9, and 10, corresponding to the Sobradinho, Três Marias, and Tucuruí datasets, respectively. In these figures, the red time series represents the actual data provided by CHESF, the company responsible for the operation of the reservoirs. The

Model	MAPE	RMSE
ARIMA	51.42	3825.54
MLP	46.10	4413.43
SVR	21.93	3162.59
ELM	27.23	3273.78
LSTM	35.76	3308.75
Static Ensemble	34.90	3282.97
Weighted Static Ensemble	31.90	3220.91
Dynamic Ensemble by KNN	40.59	3854.74
Dynamic Ensemble by Window	**18.81**	**3155.05**

Table 4: Performance Comparison of Different Models (MAPE and RMSE) on the Tucuruí Dataset

blue time series corresponds to the output of the most promising model evaluated in this study.

The x-axis represents the temporal step of the historical time series, which, in our case, is on a monthly scale. The y-axis denotes the flow rate value for the specific month associated with each point in the series.

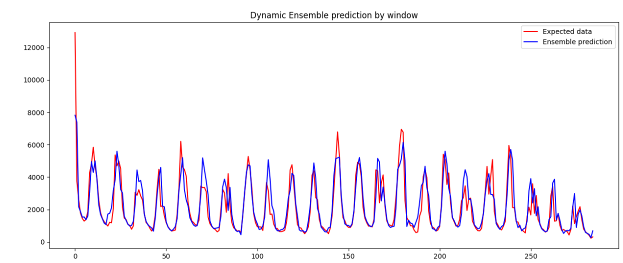


Figure 8: Predicted vs. Expected Results for the best model on the Sobradinho dataset.

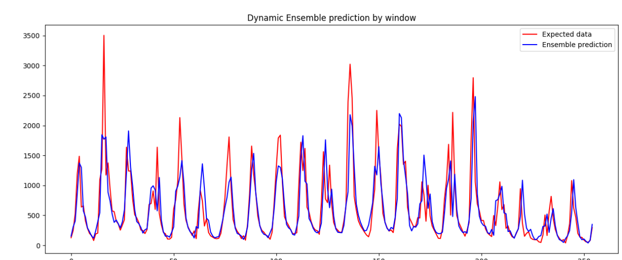


Figure 9: Predicted vs. Expected Results for the best model on the Três Marias dataset.

4 Conclusions

The first notable observation is the relatively limited performance of the ARIMA model, indicating that the dataset lacks significant linear dependencies. Nonetheless, ARIMA delivered satisfactory results in specific scenarios, warranting its occasional selection by the proposed dynamic model. This observation underscores the

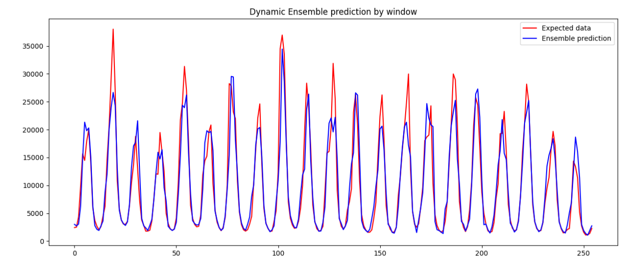


Figure 10: Predicted vs. Expected Results for the best model on the Tucuruí dataset.

approach's adaptability, effectively utilizing each model's strengths in varying contexts.

A key insight into the LSTM model's performance is its limited success on average, which highlights the inherent complexity of the forecasting task. Although LSTM produced interesting results in certain cases, the dynamic selection model's ability to replace underperforming models in real-time reinforces the practical applicability of the dynamic capabilities theory. Notably, the superior performance of the SVR model, likely attributable to its nonlinear kernel, suggests that the dataset is better suited to nonlinear modeling approaches.

In the Tucuruí dataset, shown in Figure (10), the MAPE is relatively low while the RMSE is considerably high (18.81 and 3155.05, respectively). This discrepancy is attributed to the dataset's high absolute values, necessitating the use of both metrics for a more comprehensive assessment. Small improvements in MAPE can signify substantial changes in forecast accuracy, as previously discussed. Another critical aspect is the challenge of accurately capturing peak values, which are essential in forecasting since they represent pivotal moments that can drive significant operational decisions. Among the datasets, the Sobradinho dataset most effectively captures peak values, as illustrated in Figure (8), with a MAPE of 24.79 and RMSE of 825.82, indicating a closer alignment to the true peaks. In contrast, the Três Marias dataset exhibits the highest MAPE (36.72) and an RMSE of 358.07, as shown in Figure (9), highlighting further challenges in peak prediction accuracy. Notably, the issue of peak prediction is more apparent in graphical analyses than in numerical metrics alone, emphasizing the value of visual evaluation in assessing model performance.

The superior performance of the proposed model can be attributed to the dynamic capabilities theory, which posits that no single model remains optimal throughout the forecasting process. Through iterative evaluation, the dynamic model effectively substitutes underperforming base models with those more suitable for the current conditions, yielding significant reductions in error metrics.

In contrast, static and KNN-based ensemble selection methods demonstrated less effective performance. This result suggests that certain models included in these ensemble compositions may detract from overall accuracy, a known issue that underscores the importance of careful model selection within ensemble frameworks. Dynamic selection addresses this by choosing only the most effective model at each step, streamlining convergence and enhancing efficiency.

The forecast curves generated by the proposed model show consistent and satisfactory alignment with observed data, affirming the model's suitability for real-world applications in reservoir management and similar contexts. This approach effectively balances accuracy and adaptability, making it a viable solution for dynamic and high-stakes forecasting environments.

5 Future Work

In future developments, the goal will be to optimize the model replacement timing within the ensemble. The intention is that, upon identifying that the ensemble is operating in a region unfavorable to the currently predominant model, the replacement should occur in a balanced manner. The transition should neither be so slow as to compromise the overall performance of the ensemble nor so fast as to result in a hasty and unstable pruning.

The control of this replacement speed will be governed by a coefficient called the tolerance coefficient, whose definition will be based on statistical information extracted from recent data. This approach aims to mitigate the impact of noise from past contexts and ensure that the adaptation of the ensemble is responsive to the current data conditions.

References

[1] Dihia Boulegane, Albert Bifet, and Giyyarpuram Madhusudan. 2019. Arbitrated Dynamic Ensemble with Abstaining for Time-Series Forecasting on Data Streams. In *2019 IEEE International Conference on Big Data (Big Data)*. 1040–1045. https://doi.org/10.1109/BigData47090.2019.9005541

[2] Zezhou Chen and Irena Koprinska. 2020. Ensemble Methods for Solar Power Forecasting. In *2020 International Joint Conference on Neural Networks (IJCNN)*. 1–8. https://doi.org/10.1109/IJCNN48605.2020.9206713

[3] Douglas Vieira Do Nascimento, Rafael Teixeira Sousa, Diogo Fernandes Costa Silva, Daniel Do Prado Pagotto, Clarimar José Coelho, and Arlindo Rodrigues Galvão Filho. 2023. Live Birth Forecasting in Brazillian Health Regions with Tree-based Machine Learning Models. In *2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)*. 85–90. https://doi.org/10.1109/CBMS58004.2023.00197

[4] Empresa de Pesquisa Energética (EPE). 2024. Matriz Energética e Elétrica. https://www.epe.gov.br/pt/abcdenergia/matriz-energetica-e-eletrica Accessed: November 2024.

[5] Hansika Hewamalage, Christoph Bergmeir, and Kasun Bandara. 2021. Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. *International Journal of Forecasting* 37, 1 (2021), 388–427. https://doi.org/10.1016/j.ijforecast.2020.06.008

[6] Rob J. Hyndman and George Athanasopoulos. 2018. *Forecasting: Principles and Practice* (2nd ed.). OTexts. Available at https://otexts.com/fpp3/.

[7] Haseeb Javed, Hafiz Abdul Muqeet, Amirhossein Danesh, Atiq Ur Rehman, Tahir Javed, and Amine Bermak. 2024. Impact of AI and Dynamic Ensemble Techniques in Enhancing Healthcare Services: Opportunities and Ethical Challenges. *IEEE Access* 12 (2024), 141064–141087. https://doi.org/10.1109/ACCESS.2024.3443812

[8] Bryan Lim and Stefan Zohren. 2021. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 379, 2194 (2021), 20200209. https://doi.org/10.1098/rsta.2020.0209 arXiv:https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2020.0209

[9] Diego de Souza Martins, Heder Soares Bernardino, and Luciana Conceição Dias Campos. 2024. Short-term Forecasting of the Wind Power Generation of Brazilian Power Stations Using an LSTM Model. In *Proceedings of the 20th Brazilian Symposium on Information Systems* (Juiz de Fora, Brazil) *(SBSI '24)*. Association for Computing Machinery, New York, NY, USA, Article 21, 9 pages. https://doi.org/10.1145/3658271.3658292

[10] Kasun Mendis, Manjusri Wickramasinghe, and Pasindu Marasinghe. 2024. Multivariate Time Series Forecasting: A Review. In *Proceedings of the 2024 2nd Asia Conference on Computer Vision, Image Processing and Pattern Recognition* (Xiamen, China) *(CVIPPR '24)*. Association for Computing Machinery, New York, NY, USA, Article 58, 9 pages. https://doi.org/10.1145/3663976.3664241

[11] Marcos Alberto Mochinski, Jean Paul Barddal, and Fabrício Enembreck. 2020. Improving Multiple Time Series Forecasting with Data Stream Mining Algorithms. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 1060–1067. https://doi.org/10.1109/SMC42975.2020.9283059

[12] Thiago J. M. Moura, George D. C. Cavalcanti, and Luiz S. Oliveira. 2021. On the Evaluation of Competence Measures for Time Series Forecasting. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 1527–1532. https://doi.org/10.1109/SMC52423.2021.9659086

[13] Oksana Mulesa, Anatoliy Batyuk, Fedir Geche, Olena Melnyk, Mykola Palinchak, and Tamara Radivilova. 2021. Information technology for time series forecasting based on the evolutionary method of the forecasting scheme synthesis. In *2021 IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT)*, Vol. 2. 258–261. https://doi.org/10.1109/CSIT52700.2021.9648639

[14] Bencharef Omar, Bousbaa Zineb, Aida Cortés Jofré, and Daniel González Cortés. 2018. A Comparative Study of Machine Learning Algorithms for Financial Data Prediction. In *2018 International Symposium on Advanced Electrical and Communication Technologies (ISAECT)*. 1–5. https://doi.org/10.1109/ISAECT.2018.8618774

[15] Zuokun Ouyang, Philippe Ravier, and Meryem Jabloun. 2022. Are Deep Learning Models Practically Good as Promised? A Strategic Comparison of Deep Learning Models for Time Series Forecasting. In *2022 30th European Signal Processing Conference (EUSIPCO)*. 1477–1481. https://doi.org/10.23919/EUSIPCO55093.2022.9909926

[16] Yi Peng, He Kang, and Yanan Chen. 2021. A Research on Stock Index Prediction Based on Multiple Linear Regression and ELM Neural Network. In *2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP)*. 254–259. https://doi.org/10.1109/ICSP51882.2021.9408840

[17] Eraylson G. Silva, Paulo S. G. De Mattos Neto, and George D. C. Cavalcanti. 2021. A Dynamic Predictor Selection Method Based on Recent Temporal Windows for Time Series Forecasting. *IEEE Access* 9 (2021), 108466–108479. https://doi.org/10.1109/ACCESS.2021.3101741

[18] Taiane Silva Barbosa and Artur Henrique Kronbauer. 2019. Panorama of Researches Related to the Application of Virtual Reality in the Health Area in SVR. In *2019 21st Symposium on Virtual and Augmented Reality (SVR)*. 69–76. https://doi.org/10.1109/SVR.2019.00027

[19] Simran and Jaspreet Singh. 2023. A Comprehensive Survey of PSO-ACO Optimization and Swarm Intelligence in Healthcare: Implications for Medical Image Analysis and Disease Surveillance. In *2023 3rd Asian Conference on Innovation in Technology (ASIANCON)*. 1–6. https://doi.org/10.1109/ASIANCON58793.2023.10270025

[20] Nuno Verdelho Trindade, Pedro Leitão, Daniel Gonçalves, Sérgio Oliveira, and Alfredo Ferreira. 2023. Immersive Situated Analysis of Dams' Behavior. In *2023 International Conference on Graphics and Interaction (ICGI)*. 1–8. https://doi.org/10.1109/ICGI60907.2023.10452725

[21] Hubert Truchan, Christian Kalfar, and Zahra Ahmadi. 2024. LTBoost: Boosted Hybrids of Ensemble Linear and Gradient Algorithms for the Long-term Time Series Forecasting. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management* (Boise, ID, USA) *(CIKM '24)*. Association for Computing Machinery, New York, NY, USA, 2271–2281. https://doi.org/10.1145/3627673.3679527

[22] Zheng Wang, Irena Koprinska, Alicia Troncoso, and Francisco Martínez-Álvarez. 2018. Static and Dynamic Ensembles of Neural Networks for Solar Power Forecasting. In *2018 International Joint Conference on Neural Networks (IJCNN)*. 1–8. https://doi.org/10.1109/IJCNN.2018.8489231

[23] Dong Wu, Xinyi Lin, Shivam Gupta, and Arpan Kumar Kar. 2024. Big Data Analytics Capability, Dynamic Capability, and Firm Performance: The Moderating Effect of IT–Business Strategic Alignment. *IEEE Transactions on Engineering Management* 71 (2024), 11638–11651. https://doi.org/10.1109/TEM.2024.3429648

[24] Songting Xing and Yuansheng Lou. 2019. Hydrological time series forecast by ARIMA+PSO-RBF combined model based on wavelet transform. In *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*. 1711–1715. https://doi.org/10.1109/ITNEC.2019.8729367

[25] Shuhan Zhong, Sizhe Song, Weipeng Zhuo, Guanyao Li, Yang Liu, and S.-H. Gary Chan. 2024. A Multi-Scale Decomposition MLP-Mixer for Time Series Analysis. *Proc. VLDB Endow.* 17, 7 (May 2024), 1723–1736. https://doi.org/10.14778/3654621.3654637

Received 21 October 2024; revised 03 February 2025; accepted 11 March 2025