A Comparative Analysis of Combination Operators in Heterogeneous Ensembles for One-Step-Ahead Time Series Forecasting

Rodolfo Viegas de Albuquerque[1], **João Fausto Lorenzato de Oliveira**[1]

[11]Programa de Pós-Graduação em Engenharia da Computação (PPGEC),
Universidade de Pernambuco (UPE) – Recife – PE – Brazil

rva@ecomp.poli.br, fausto.lorenzato@upe.br

Abstract. *This study investigates the efficacy of four combination operators in heterogeneous ensembles for one-step-ahead time series forecasting tasks: simple mean, median, stacking with Support Vector Regression (SVR) as a meta-model, and weighted average. Nine machine learning and statistical models were trained and subsequently their outcomes were averaged. The results show that the simple mean operator outperforms the median operator in terms of RMSE and SMAPE, yielding an average error of 15.888% and 17.791%, respectively, when compared to individual models and previous iterations of themselves. In contrast, the novel weighting average operator yielded the lowest average error values for both metrics, indicating potential for further enhancements and research. In addition to the increase in accuracy, the data analysis reveals that for monthly time series and within a forecasting framework, there is a tendency for the standard deviation to decrease as more models are incorporated.*

1. Introduction

The task of forecasting time series represents a significant domain within computational intelligence, with diverse applications across fields such as economics, and social sciences. These data are characterized as sequences of observations recorded sequentially over time [Box et al. 2015], and through the analysis and forecasting of time series data, it is possible to provide essential insights for decision-making processes in various areas.

In general, the combination of models has the potential to surpass the efficacy of individual models. According to [Clemen 1989], the use of combination leads to an increase in the accuracy of forecasting tasks. Various combinatory approaches have been applied. [Makridakis and Winkler 1983] demonstrates the effectiveness of a simple average by integrating 14 methods and evaluating them on 1001 series from the M1-Competition. The arithmetic mean has notably achieved significant successes, as evidenced by the M5 Competition, where the top three positions employed the simple mean [Makridakis et al. 2022].

The study conducted by [Petropoulos and Svetunkov 2020] presented a model employing the median operator. In contrast, the exploration in [Kourentzes et al. 2014] validated the median operator's potential, as it produced significant results by aggregating multiple Multi-Layer Perceptrons (MLPs). To evaluate the effectiveness of Deep Learning (DL) models in executing time series forecasting, [Makridakis et al. 2023] utilized

the median operator to amalgamate four diverse DL model categories, each comprising 50 models, resulting in an ensemble integrating all 200 models.

The exploration of weighted averages has been extensively addressed in numerous studies. In particular, [Granger and Ramanathan 1984] utilized linear regression coefficients as weights, while [Conflitti et al. 2015] used LASSO regression coefficients. [Montero-Manso et al. 2020] represents a state-of-the-art effort whereby the author formulated a weighted combination employing the XGBoost model to estimate weights derived from time series features.

The stacking methodology has been utilized in forecasting tasks [Ribeiro and dos Santos Coelho 2020]. Additionally, hybrid systems such as [de Oliveira et al. 2021, De Oliveira and Ludermir 2016] integrate linear and non-linear decomposed predictions through summation.

This research investigates the efficacy of four combination operators: the simple mean, median, stacking with SVR as a meta-model, and an innovative operator referred to as weighted average by feature importance (WAFI). Feature importance is derived from the Extremely Randomized Trees model. The study employs 30 monthly time series from the M3 Competition [Makridakis and Hibon 2000], applying nine methods that are subsequently combined using the four operators to generate one-step-ahead forecasts. The ensuing testing errors are assessed and compared internally and against individual models.

2. Weighted Average by Feature Importance

This study proposes a novel method: a linear combination through a weighted average for one-step-ahead time series forecasting. It employs a tree-based model, specifically Extremely Randomized Trees, to calculate the weights for the weighted mean. From these tree-based models, feature importance is determined and utilized as a significance factor for the mean operator, thereby integrating the predictions of individual models.

2.1. Simple Average and Weighted Average

A commonly utilized approach for combining forecasts is the simple average (SA). In this method, all instances x_i are assigned equal weights, which implies that all data are considered equally significant. Each of these instances is summed and subsequently divided by its total quantity. Within the context of time series, a SA is expressed by x_i as the forecast of the i-th model from a collection of T methods, after which these forecasts are divided by the total number T of models utilized, expressed in 1:

$$average = \frac{1}{T}\sum_{i=1}^{T} x_i \tag{1}$$

The SA is a special case of weighted average (WA), i.e., when all instances have the same importance, which are represented by weights w_i that multiply all entries x_i. The equation 2 represents the weighted average:

$$weighted = \sum_{i=1}^{T} w_i x_i \tag{2}$$

A method to estimate the weights for a WA operator is through the application of linear regression (LR), as stated by [Kuncheva 2014]. LR 3 is a conventional statistical approach used to model relationships among variables. It is based on certain assumptions, including the normal distribution of residuals, constant variance (homoscedasticity), independence of residuals (absence of autocorrelation), and the absence or low level of correlation among independent variables (multicollinearity).

$$f(x) = \beta_0 + \beta_1 x_1 + ...\beta_n x_n \tag{3}$$

2.2. Feature Importance

The fundamental aspect of the proposed combination system lies in its ability to estimate weights for computing a WA combination. This capability is facilitated by employing the technique known as feature importance (FI) in decision tree-based models. Decision trees represent a machine learning (ML) method that, unlike LR, can address variables that exhibit non-linear relationships with the target variable and among themselves, such as multicollinearity [Molnar 2025]. Moreover, decision trees can be highly interpretable models, capable of illustrating relationships between variables and the target by either plotting the tree after fitting or utilizing FI to indicate the relative explicability of different variables.

The methodology for determining FI, as articulated by [Murphy 2022], quantifies the significance of a variable k within a decision tree T by implementing the formula presented in $R_k(T)$, 4:

$$R_k(T) = \sum_{j=1}^{J-1} G_j \mathbb{I}(v_j = k) \tag{4}$$

The aggregate of information gain G_j — which may be defined as the reduction of mean squared error (MSE) in the context of regression trees, or as the decrease in impurity for classification trees — is computed for each non-leaf node j representing a particular feature k. The indicator function $\mathbb{I}(v_j = k)$ assigns a value of 1 if the j node encodes the feature k, and 0 otherwise. Subsequently, the resulting measure of FI may be normalized to unity.

In ensemble methods, such as Random Forest (RF), the computation of FI is derived by averaging across all internal trees for the given variable. It is represented by the equation 5

$$R_k = \frac{1}{M} \sum_{m=1}^{M} R_k(T_m) \tag{5}$$

Upon the completion of the process, a feature importance vector was generated.

2.3. Extremely Randomized Trees

The Extremely Randomized Trees (ET), as introduced in [Geurts et al. 2006], is an ensemble method designed to mitigate the high variance typically associated with Decision

Tree models. The primary distinction between RF and ET lies in their data splitting techniques; RF employs the Bootstrap method to generate subsets of observations with replacement, maintaining the original size of the training set, and subsequently selects features randomly from these subsets to construct trees. In contrast, ET utilizes the entire set of observations, randomly selecting features and thresholds for node splitting. Some advantages in using ET instead of other tree-based models are: ET is faster than RF, which contributes to experiments, and it also has lower variance that diminishes the possibility of over-fitting, and the selecting the best features to do better generalization (it is possible to return a suitable FI vector which may be a more appropriate weights). Moreover, ET is robust for noisy data[Ghazwani and Begum 2023], which is an important feature that is present in time series data.

3. Methodology

This section delineates the protocol for training and evaluating the models, including their ensemble. The system is designed such that the training and testing phases are executed sequentially within a single phase.

3.1. System

The initial phase of the system involves the acquisition and segmentation of time series for subsequent utilization. Specifically, the first 30 monthly time series from the M3 Competition dataset are procured and divided into two distinct datasets: training and testing. The testing dataset comprises 18 observations, consistent with the M3 Competition specifications and in adherence to the methodology put forth by [Makridakis and Winkler 1983] and by recommendation in [Hyndman and Athanasopoulos 2021]. All the time series have a fixed number of 12 lags for ML models without other preprocessing methods.

After the acquisition of the series, the training phase commences, wherein five statistical models (ARIMA, ETS, CES, Theta, and TBATS) are applied to the training dataset. Subsequently, four ML models, namely MLP, KNN, SVR, and RF, are trained. Upon fitting both statistical and ML models, the system computes their respective fitted values (training predictions) and the one-step-ahead test prediction.

The primary objective of this study is to evaluate four categories of operators utilizing various ensemble methods, subsequently averaging the outcomes. To achieve this, it is imperative to construct all possible combinations, without repetition, ranging from two methods up to nine (the total number of models analyzed in this study). The methodology for executing this proposal is articulated in 1:

The aforementioned algorithm 1 elucidates the procedure for constructing method combinations. Initially, the outer loop iterates across the set of 30 time series, during which each of the nine individual models is trained. Subsequently, the internal loop sequentially processes each method trained on the respective series. From 9 to 2 methods, all conceivable combinations, as delineated in equation 6, are formulated. For each specific combination, the four operators—mean, median, weighted average based on FI, and SVR-Stacking—are employed to generate forecasts.

For each set of 30 time series, 511 distinct types of combinations are generated, encompassing between 2 and 9 models. Each combination employs four operators. In total, 61,320 ensemble configurations were constructed.

Algorithm 1 Combinations

for $i \leftarrow 1$ to 30 **do**
    $Predictions \leftarrow Training(StatiscalModels, Y^i_{train})$
    $Predictions \leftarrow Training(MLModels, Y^i_{train})$
    for $j \leftarrow 9$ to 2 **do**
        $Combinations \leftarrow C^9_j$
        $Operators(Combinations, Predictions)$
    end for
end for

Table 1. Hyper-parameters Table.

Model	Hyper-parameters
MLP	hidden units: 20, 50, 100 learning rate: np.logspace(-5, -1, 15)
KNN	number of neighbors: range(1,15)
SVR	C: 10, 100, 1000, gamma: 0.1, 0.01, 0.001, epsilon: 0.1, 0.01, 0.001
Random Forest	number of estimators: 50, 100, 200 max depth : 5, 10 max features: 0.6, 0.8, 1,
Extremely Randomized Trees	number of estimators: 100 max depth: no limit max features: 1

$$C^9_j = \frac{9!}{j!(9-j)!} \tag{6}$$

The diagram1 describes the entire process:

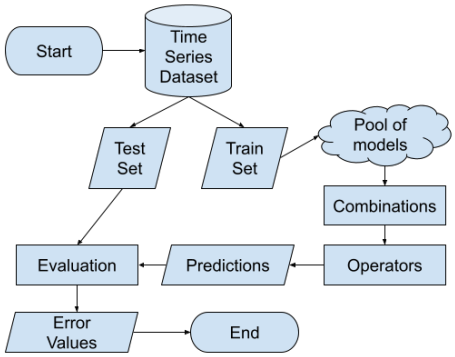


Figure 1. Diagram of the System.

3.2. Hyper-parameter Optimization

The nine individual models were each fitted once across all 30 time series. The ML techniques were trained using Random Search with 30 iterations over a range of hyper-parameters. This particular optimization method was selected due to its superior speed compared to Grid Search and its enhanced efficiency in identifying the optimal set of hyper-parameter values. The parameters requiring optimization are detailed in 1.:

The SVR-Stacking operator employs the identical set of hyperparameters utilized in the standalone SVR methodology, accompanied by Random Search executed over 30 iterations. Due to constraints in computational resources, the hyperparameters for the Extremely Randomized Trees used in the weighted average were not optimized. The implemented system is accessible on GitHub[1]

3.3. Metrics

This study analyzed the results using the metrics RMSE and SMAPE. [Makridakis and Winkler 1983] utilized only one metric, namely MAPE. Using a single metric for data analysis raises concerns regarding potential bias, as it provides a singular perspective on the data. Furthermore, the metric MAPE, despite its popularity in evaluating time series forecasting problems, presents an inherent limitation: it cannot handle zero values due to the impossibility of division by zero when the actual value is zero. When the denominator is near zero, results may become unstable; additionally, this metric disproportionately penalizes negative errors more than positive ones. The SMAPE metric addresses this issue. The formula for the metric is provided in 7.:

$$SMAPE = \frac{100\%}{N} \sum_{i=0}^{N-1} \frac{2 * |y_i - \hat{y}_i|}{|y| + |\hat{y}|}. \tag{7}$$

The additional metric employed is the Root Mean Square Error (RMSE), which, although scale-dependent, offers the benefit of being readily interpretable. Furthermore, it provides an alternative perspective that helps to mitigate the bias associated with relying solely on a single forecasting error metric. The formula for RMSE is provided in 8 :

$$RMSE = \sqrt{\frac{\sum_{i=0}^{N-1}(y_i - \hat{y}_i)^2}{N}}. \tag{8}$$

Upon the completion of all experiments, the RMSE and SMAPE metric results were aggregated by their means for the purpose of constructing charts displaying RMSE/SMAPE values as a function of the number of methods in the ensembles. Each line graph comprises four curves, each representing one of the four operators.

4. General Analysis and Results

This section presents the analysis of the average RMSE and SMAPE values as a function of the number of models in the ensemble for all grouped series. The values are displayed in Tables 2 and 3.

[1]https://github.com/RodolfoViegas/data_analysis_of_four_operators

Table 2. Average RMSE Values as a function of the number of models in the ensemble.

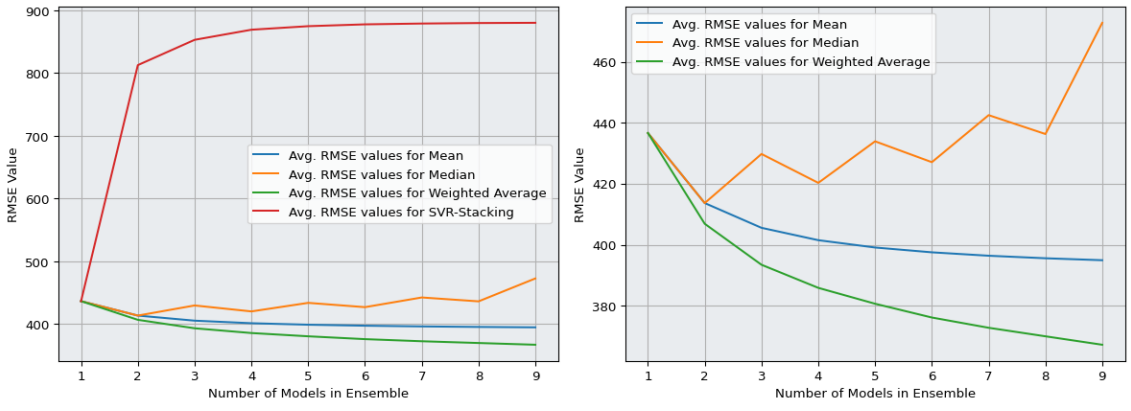
Nº Models	Simple Mean	Median	WAFI	SVR-Stacking
2	413.704	413.704	406.92	812.363
3	405.592	429.793	393.485	852.605
4	401.542	420.327	385.919	868.687
5	399.143	433.941	380.714	874.316
6	397.564	427.107	376.198	877.221
7	396.448	442.518	372.836	878.519
8	395.619	436.335	370.042	879.328
9	394.979	472.767	**367.276**	879.686

The grouping of RMSE and SMAPE values according to the mean error value across the number of combined models (considering all time series) demonstrates that both the arithmetic mean and the weighted mean, adjusted by FI, surpass the performance of the other two combiners. The technique of stacking — utilizing the nonlinear SVR meta-model — exhibits a marked deterioration in performance with two combined methods when compared to the individual model values and this decline persists until the number of methods reaches six, at which point it stabilizes. The median achieves its optimal performance with two methods in the ensemble, paralleling the arithmetic mean, yet exhibits fluctuations that marginally degrade its performance. These error value behaviors pertain to both metrics.

Table 3. Average SMAPE Values as function of number of models in ensemble.

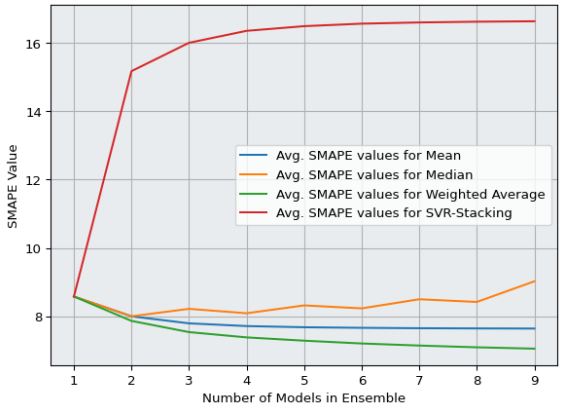
Nº Models	Simple Mean	Median	WAFI	SVR-Stacking
2	7.998	7.998	7.863	15.177
3	7.792	8.217	7.535	16.007
4	7.712	8.087	7.380	16.358
5	7.678	8.317	7.284	16.494
6	7.661	8.230	7.202	16.568
7	7.649	8.497	7.142	16.603
8	7.644	8.418	7.090	16.625
9	7.640	9.023	**7.051**	16.638

The optimal average RMSE value achieved through the simple mean method is 394.979, based on the combination of nine methods. In contrast, the average RMSE value for individual models is 436.653, as referenced in 9. This signifies an improvement percentage of 9.543%. Regarding the SMAPE metric, the average error for individual models is 8.577, whereas the simple mean approach yields an average error of 7.64. This represents an improvement in the error metric by 10.924%. When considering the WAFI method in comparison to individual models, the enhancements in the average RMSE and SMAPE values amount to 367.276 and 7.051, respectively, equating to improvements of 15.888% and 17.791%.

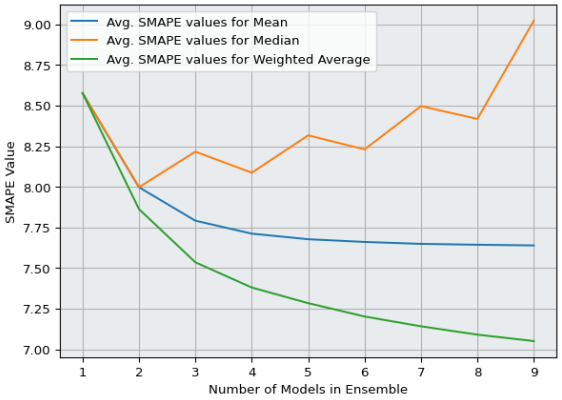


(a) RMSE with SVR-Stacking

(b) RMSE without SVR-Stacking



(c) SMAPE with SVR-Stacking



(d) SMAPE without SVR-Stacking

Figure 2. Average RMSE and SMAPE values as function of number of models for all grouped series.

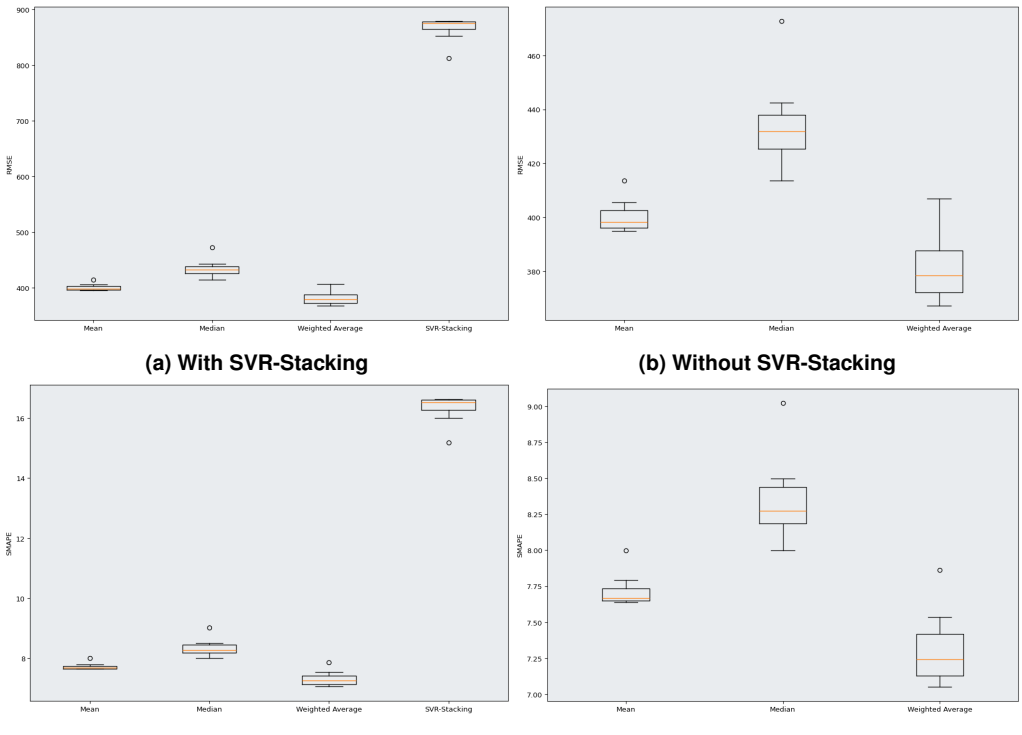
The performance of the SVR-Stacking and the median operator was inferior, not only when compared to individual models but also to previous iterations of themselves. As additional models were incorporated into the ensemble, the accuracy of the combined model decreased. The most accurate result in stacking was achieved with two methods, resulting in an RMSE of 812.363 and an SMAPE of 15.177. On average, the stacking operator exhibited overfitting, leading to a deterioration of 86.043% in terms of RMSE and 76.95% in SMAPE. A potential explanation for these outcomes is the chaotic behavior [Makridakis et al. 1998] that non-linear models may exhibit when fitting non-linear series. Chaos is a characteristic in time series where some observations appear to behave randomly, even though they follow a deterministic process. The non-linear nature of the SVR meta-model employed may exacerbate this aspect; thus, including more models in the stack resulted in chaotic outcomes. In addition [Wang et al. 2023] highlight that the stacking method is inefficient in using the training set and to improve its performance they recommend cross-learning, which is training using many series instead of one to extract information.

The median operator had also an average behavior which adding more models the more worse was the results; however, the decline of accuracy was less worse than SVR-Stacking.

$$PercentualImprovement = \frac{Error1 - Error2}{Error1} * 100. \quad (9)$$

4.1. Box plot and Dispersion Analysis

The box plots illustrating the average RMSE and SMAPE values 3 for the combination operators provide a more detailed view of the distribution of error data. SVR-Stacking's box plot lies above those of all other operators, indicating a lack of convergence in distribution, which might suggest similar mean error values. However, this is not the case as the optimal value for stacking is an outlier. Further analysis of the remaining data can be conducted after excluding the stacking box plot. There is no overlap among the boxes, nor do the median lines of each box lie outside the bounds of the others. This evidence suggests that the three groups are distinct; specifically, the accuracies of the median, simple, and WAFI methods are not equivalent, with WAFI demonstrating superior performance according to both metrics.



(a) With SVR-Stacking **(b) Without SVR-Stacking**

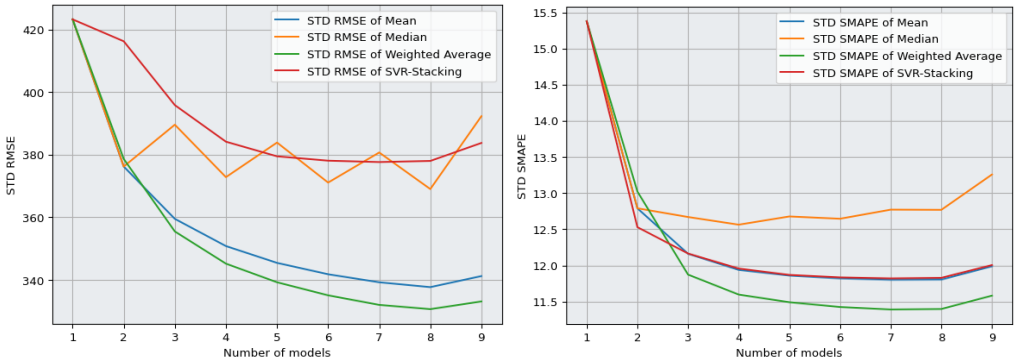
(c) With SVR-Stacking **(d) Without SVR-Stacking**

Figure 3. Boxplots of average RMSE and SMAPE

A salient feature observable from the graphs 4 is that as the number of models integrated into the combination increases, the standard deviation of the mean error values, RMSE and SMAPE, decreases. Variance and standard deviation serve as dispersion metrics, reflecting a degree of robustness in the models.

Employing the analogy of a target and darts [Domingos 2012], models exhibiting low variance or standard deviation generally achieve higher accuracy at specific points. Models with diminished dispersion present decreased uncertainty regarding accuracy, rendering the standard deviation a crucial metric for assessing probability distributions.

When residuals are normally distributed, uncorrelated, and homoscedastic, estimating the probability distribution of the forecasts becomes feasible.



(a) RMSE **(b) SMAPE**

Figure 4. Average SMAPE values as function of number of models for all grouped series

The Friedman test and the Nemenyi test, as referenced in *post hoc*, were conducted on a dataset encompassing average error values ranging from 2 to 9 combinations across all 30 time series, with a significance level of 0.05. The Friedman test did not reveal any significant differences among the operators; nevertheless, as indicated in *i.e.*, the null hypothesis could still be rejected. Subsequently, the Nemenyi test was executed to ascertain the ranking of four operators. Figure 5 presents the Critical Distance Diagram [Demšar 2006], illustrating the RMSE and SMAPE metrics. The proposed method achieved the highest rank, followed by the SA.

(a) RMSE **(b) SMAPE**

Figure 5. Critical Distance Diagrams for RMSE and SMAPE metrics.

5. Conclusion

Following the analysis of two datasets, each containing 15,331 lines, evidence was identified corroborating the well-documented assertion that combination models can surpass single model performance [Clemen 1989]. Among the four operators examined, the simple mean yielded promising results, ranking as the second most accurate operator, in accordance with early findings in [Makridakis and Winkler 1983].

In addition to the increase in accuracy, the data analysis reveals that for monthly time series and within a one-step ahead forecasting framework, there is a tendency for the standard deviation to decrease as more models are incorporated, which is also in agreement with [Makridakis and Winkler 1983] who found the decrease of MAPE variance when the number of methods increases in the ensembles. This attribute can be particularly critical in fields such as finance, which assess decision-making risks.

Within the scope of this study, focusing on one-step-ahead forecasting and monthly time series analysis, certain characteristics have been identified. The median operator, which has demonstrated some evidence of good performance [Kourentzes et al. 2014], did not exhibit satisfactory results in the general analysis; its performance was inconsistent as more models were incorporated into the ensemble. The Stacking method, employing SVR as the meta-model, yielded the poorest results of the entire study, exhibiting overfitting with merely two stacked models and deteriorating further with adding more methods. These findings pertain to the general analysis, where all series are aggregated, and are not intended to be generalized to other contexts.

A significant observation was the superior performance attained by the weighted average derived from the feature importance of Extremely Randomized Trees. The primary hypothesis positing this outcome suggests that the feature importance effectively extracts underlying patterns from the data. These numerical representations can serve as weights by numerically representing the relevance of features, specifically, the forecasts of models. Another important fact is the ET model wasn't optimized in its parameters, which shows the computational efficiency of the WAFI system and the default parameters could assign importance to single models.

In future research, evaluating the accuracy of the four operators in multi-step forecasting is warranted, employing time series that exhibit various seasonal patterns. The aim is to ascertain whether these operators can sustain stability when forecasting horizons are extended and uncertainties are prevalent. Furthermore, exploring alternative tree-based ensembles, such as Random Forest, is advisable to assess their performance compared to the mean, median, and ET, determining if they can achieve superior results. Another point to take into consideration for future work is the hyperparameter optimization on the WAFI system, that is, to test if optimizing its parameters could improve the significance of FI vector for combining. Additionally, hybrid models may be developed by applying the WAFI operator to model the linear or non-linear components of the series effectively. And comparing the WAFI against state-of-the-art models, like FFORMA, to analyse if its performance may surpass them.

References

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International journal of forecasting*, 5(4):559–583.

Conflitti, C., De Mol, C., and Giannone, D. (2015). Optimal combination of survey forecasts. *International Journal of Forecasting*, 31(4):1096–1103.

De Oliveira, J. F. and Ludermir, T. B. (2016). A hybrid evolutionary decomposition system for time series forecasting. *Neurocomputing*, 180:27–34.

de Oliveira, J. F., Silva, E. G., and de Mattos Neto, P. S. (2021). A hybrid system based on dynamic selection for time series forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, 33(8):3251–3263.

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine learning research*, 7:1–30.

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10):78–87.

Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, 63:3–42.

Ghazwani, M. and Begum, M. Y. (2023). Computational intelligence modeling of hyoscine drug solubility and solvent density in supercritical processing: Gradient boosting, extra trees, and random forest models. *Scientific Reports*, 13(1):10046.

Granger, C. W. and Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of forecasting*, 3(2):197–204.

Hyndman, R. and Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. OTexts, 3 edition.

Kourentzes, N., Barrow, D. K., and Crone, S. F. (2014). Neural network ensemble operators for time series forecasting. *Expert Systems with Applications*, 41(9):4235–4244.

Kuncheva, L. I. (2014). *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons.

Makridakis, S. and Hibon, M. (2000). The m3-competition: results, conclusions and implications. *International journal of forecasting*, 16(4):451–476.

Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, 38(4):1346–1364. Special Issue: M5 competition.

Makridakis, S., Spiliotis, E., Assimakopoulos, V., Semenoglou, A.-A., Mulder, G., and and, K. N. (2023). Statistical, machine learning and deep learning forecasting methods: Comparisons and ways forward. *Journal of the Operational Research Society*, 74(3):840–859.

Makridakis, S., Wheelwright, S., and Hyndman, R. J. (1998). *Forecasting: methods and applications*. John Wiley & Sons.

Makridakis, S. and Winkler, R. L. (1983). Averages of forecasts: Some empirical results. *Management science*, 29(9):987–996.

Molnar, C. (2025). *Interpretable Machine Learning*. 3 edition.

Montero-Manso, P., Athanasopoulos, G., Hyndman, R. J., and Talagala, T. S. (2020). Fforma: Feature-based forecast model averaging. *International Journal of Forecasting*, 36(1):86–92.

Murphy, K. P. (2022). *Probabilistic Machine Learning: An introduction*. MIT Press.

Petropoulos, F. and Svetunkov, I. (2020). A simple combination of univariate models. *International journal of forecasting*, 36(1):110–115.

Ribeiro, M. H. D. M. and dos Santos Coelho, L. (2020). Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Applied soft computing*, 86:105837.

Wang, X., Hyndman, R. J., Li, F., and Kang, Y. (2023). Forecast combinations: An over 50-year review. *International Journal of Forecasting*, 39(4):1518–1547.