Vol. 34, No. 3 (2024) | DOI: 10.37190/ord240306

OPEN ACCESS

Operations Research and Decisions

www.ord.pwr.edu.pl

OPERATIONS RESEARCH AND DECISIONS QUARTERLY

# Does a meta-combining method lead to more accurate forecasts in the decision-making process?

Emrah Gulay[1*] Serkan Aras[1]

[1] *Department of Econometrics, Dokuz Eylul University, Turkey*
[*] *Corresponding author, email address: emrah.gulay@deu.edu.tr*

**Abstract**

To improve forecasting accuracy, researchers employed various combination techniques for a long time. When researchers deal with time series data by using dissimilar models, the combined forecasts of these models are expected to be superior. Deriving a weighting scheme performing better than simple but hard–to–beat combining methods has always been challenging. In this study, a new weighting method based on the hybridisation of combining algorithms is proposed. Five popular datasets were utilised to demonstrate the effectiveness of the proposed method in an out-of-sample context. The results indicate that the proposed method leads to more accurate forecasts than other combining techniques used in the study.

**Keywords:** *forecasting, combined forecasts, time series, meta–combining*

## 1. Introduction

In contrast to developments in individual forecasting algorithms in the field of predictive analytics, a combining algorithm is still the leading area of focus as well as a fundamental approach for improved accuracy. Predictive analytics of time series data in forecasting plays an important role in the process of extracting meaningful information from historical data. Similar to many other predictive problems, forecasting a time series requires a series of expert interventions starting with diagnostic tests and ending with the production of hold-out forecasts. Therefore, traditional forecasting techniques need some implicit assumptions including advanced knowledge of methodology and terminology. In this context, the effectiveness of predictive analysis is ultimately influenced by the level of expertise in the conventional forecasting process. On the other hand, the need for expert intervention is a drawback arising from the bias produced in forecasting due to inappropriate implementation and intentional manipulations through the improper use of predictions. It is critical to develop a weighting scheme that combines individual models and eliminates the user's need for expert guidance on which model to choose.

Received 2 June 2023, accepted 2 July 2024, published online 17 October 2024
ISSN 2391-6060 (Online)/© 2024 Authors

The costs of publishing this issue have been co-financed by the program *Development of Academic Journals* of the Polish Ministry of Education and Science under agreement RCN/SP/0241/2021/1

It is generally acknowledged that combining several different predictive models has a predictive advantage in time series forecasting. In the field of forecasting time series data, hybrid models have attracted much attention for a long time [8, 23]. In recent years, meta-combining methods such as a combination of combining algorithms have become more efficient and have exhibited superior performance in terms of forecast accuracy [12, 19, 34]. The rationale behind a meta-combining approach is to cope with controversy when the results of single forecasting models differ and combining algorithms make the decision-making process more challenging in terms of forming different weighting schemes.

In this paper, a novel combining algorithm known as meta-combining is proposed to neutralise the subjectivity in deciding the weight of any single model by hybridising weighting algorithms and adjusting the rank on the two popular combining methods. To guarantee the principal properties of the proposed weighting algorithm, the provision of highly unbiased accuracy forecasts and better forecast direction accuracy has been considered in light of the controvertible facts in the forecasting rule.

The novel combining approach for the time series forecasting task that is presented in the study is based on the compression of two important forecasting rules into a single decision procedure. To the best of our knowledge, this is the first study to combine methods in one equation by considering two substantial forecasting procedures in the time series forecasting application. An extensive set of datasets and benchmark models confirm that the proposed algorithm leads to better performances. We believe that this study can have a considerable impact on the field of forecasting time series data.

The motivation of this study is twofold: on the one hand, according to the study in [34], complex models tend to provide better fits to in-sample data, but simple models are often assumed to be more accurate when predicting the future. This dilemma presents a significant challenge for decision-makers who seek reliable and accurate forecasts for future outcomes. Complex models, with their ability to capture intricate relationships within the data, might lead to overfitting and poor generalization beyond the observed data, resulting in less accurate predictions. On the other hand, simple models, while less prone to overfitting, might overlook important patterns and lead to less precise forecasts. Striking a balance between these two contrasting aspects is crucial to achieving accurate and robust predictions.

To address this challenge, researchers have turned their attention to hybrid algorithms for combining forecasting methods. These hybrid approaches leverage the strengths of both simple and complex models, aiming to create a more robust and accurate forecasting framework. By effectively combining multiple decision rules into a single rule, these hybrid algorithms have shown promising results in various domains. Furthermore, the success of hybrid algorithms hinges on the use of an appropriate and robust weighting scheme. Two such weighting schemes that have gained popularity are the forecast direction scheme and the inverse rank weighting scheme.

The forecast direction weighting scheme evaluates the past performance of each individual forecasting method and grants greater importance to those that consistently make correct directional predictions. In other words, it assigns higher weights to methods that accurately anticipate whether the future value will be higher or lower than the current one. This strategy is based on the notion that forecasting methods with a history of reliable directional predictions are more likely to provide valuable insights to the hybrid model, thereby bolstering its overall predictive accuracy. By emphasizing the contributions of such well-performing methods, the forecast direction weighting scheme aims to enhance the hybrid model's ability to make accurate predictions.

On the contrary, the inverse rank weighting scheme centers on the comparative performance of each forecasting method in relation to others. It allocates greater weights to methods that have demonstrated superior historical performance when ranked against alternative methods. This approach ensures that the most successful and competitive methods exert a stronger influence in the hybrid model, while de-emphasizing methods that have exhibited lower consistency or accuracy over time. By leveraging the rankings of forecasting methods, the inverse rank weighting scheme seeks to bolster the hybrid model by giving prominence to the most effective and reliable approaches in the ensemble.

By integrating both the forecast direction and inverse rank weighting schemes, a comprehensive assessment of each forecasting method's performance is achieved, both on an individual basis and in comparison, to others. This inclusive approach facilitates the development of a robust hybrid forecasting model that can adeptly adapt to the intrinsic complexities present in the data. The hybrid model generated through this methodology is capable of delivering accurate predictions while effectively mitigating the risks associated with overfitting or underperformance. In essence, the combination of these weighting schemes enhances the model's overall predictive capability, making it a valuable tool for reliable and informed decision-making.

In conclusion, the primary aim of this study is to address the trade-off between the strengths of complex and simple forecasting models by harnessing the potential of hybrid algorithms. By fusing the best aspects of both approaches, the hybrid model aims to achieve a balance that yields accurate predictions while avoiding the pitfalls of overfitting and limited generalization.

Furthermore, the incorporation of robust weighting schemes, such as the forecast direction and inverse rank weighting schemes, plays a pivotal role in enhancing the hybrid model's accuracy and reliability. The forecast direction scheme leverages historical performance to prioritize methods with consistent and accurate directional predictions, while the inverse rank weighting scheme considers relative performance to elevate the most successful methods and diminish less reliable ones. This comprehensive approach ensures that the hybrid model can adeptly adapt to different forecasting scenarios, making it a valuable tool for decision-makers across a diverse range of applications.

The paper is organised as follows. Section 1 provides an introduction. Section 2 reviews the literature, followed by Section 3, which presents the datasets and methodology. The forecasting results and discussion are provided in Section 4. Section 5 provides the conclusion and future directions.

## 2. Literature

The literature review highlighted the importance of combining methods to obtain better forecasts than individual models. While the literature provided useful guidance overall on the shortcomings of the widely used combining methods, it also served to highlight the selection of the combining methods in terms of the problem at hand. Combining forecasts of various competing models to increase accuracy dates back to the 1960s. Papers [42] and [7] shed light on improving forecast performance using multiple methods effectively as an alternative to trusting only a single method. Besides these seminal studies, the literature comprises many papers that present the idea of combined forecasts in quite different application areas. Some excellent papers [14, 16, 52] provide a snapshot review of the literature up to the time they were published. Table 1 presents a systematic overview of the studies in the field of combining methods. More recent papers have put the idea of combined forecasts into practice in diverse fields such as solar power forecasting [18], coronavirus disease forecasting [17], electricity price forecasting [37], air pollutant concentration [55], volatility forecasting in portfolio selection [11], wind speed forecasting citejiang2021combined, and forecasting seasonal influenza outbreaks [54].

**Table 1.** The systematic overview of the related literature

| Ref. | Single models | Combining method(s) | Key findings |
|---|---|---|---|
| [41] | 15 individual regression models deriving from 15 variables | mean, median, trimmed mean, and combined method based on the combining weights | compared to individual regression models, the forecast combination method reduced the volatility in the forecast significantly by including information from various economic variables |
| [45] | autoregressive model, 73 recursively produced regression models based on individual predictors for forecasting the output growth of seven countries | time-varying-parameter combination, factor models, shrinkage models, weighted combination based on historical individual performance, and simple combination | the simple mean and the trimmed mean produced the lowest squared error loss; the sophisticated combination methods performed worse than the simple ones |
| [57] | autoregressive integrated moving average model (ARIMA), artificial neural networks (ANN) | the equal weights method | the obtained results indicated that the neural network model outperformed ARIMA in all error measurements but performed worse than the Mean combining method |
| [13] | restricted model, unrestricted model, ridge regression | Bayesian model averaging, the weighted combination based on individual models, equal weights, the proposed approach of combining forecasts from nested models | the proposed method gained an advantage over ridge regression and Bayesian model averaging but performed equally well with equal weights |
| [24] | 14 forecasting methods from the M3-competition [36] | The simple Mean combination of the proposed simple model--selection criterion | even the worst-performing combining method yields significantly better results than the worst-performing single method; therefore, opting for a combining method carries a lower risk than selecting a single model |
| [31] | 22 forecasting methods from the M3-competition [36] | winsorised and trimmed means were used as more robust combining methods | the results indicated that trimmed and Winsorised means generate more accurate forecasts especially in the case of high variance in the forecasts of the single models |
| [21] | the expert forecasts of the ECB survey of professional forecasters | the combining methods relying on principal components, trimmed means, simple mean, optimal weights obtained by least square estimates, and Bayesian shrinkage | it is very hard to beat the simple mean and it is not possible to determine the best combining method for all variables and horizons considered |

**Table 1.** The systematic overview of the related literature

| | | | |
|---|---|---|---|
| [3] | ARIMA, self-exciting threshold autoregressive model (SETAR), logistic smooth transition autoregressive model (LSTAR), ANN, least support vector machines (LS-SVM) | mean, median, trimmed means, and the proposed combination method based on the mean and median combination | the proposed method outperformed the combining and single techniques on the six real-world time series |
| [33] | four groups of different forecasting methods: simple forecasting models, automatic Box—Jenkins models, structural time series models, and computational intelligence models | simple average, simple average with trimming, the variance-based model, the outperformance method, the variance-based pooling | setting the weights of a combination depending on its ranking results in a superior performance over single model selection techniques |
| [22] | poll projections, expert judgment, quantitative models, the forecasts of the Iowa Electronic Markets | the equal weights method | the performance increased when a simple combining method was used, especially when the number of single forecasts became larger and the single models conveyed as much varied information as possible |
| [27] | regression models relying on different variables | the simple mean, the optimal weighted combination based on the past performance, the combination method based on principal components | they found that, among other weighting schemes, combined forecasts with a simple average exhibit better out-of-sample performance than combining methods based on principal components |
| [9] | ARMAX, linear regression, Markov regime switching model, and time-varying regression | different versions of the equal weighted combination method | employing simple averages among more sophisticated methods leads to significantly better performances than single models in 33% of all cases examined |

# 3. Data and methods

## 3.1. Datasets

In this section, the datasets utilised for the analysis are introduced. To highlight the superiority of the proposed method, five datasets are used to compare the method's performance with five single models and six combining methods in the scope of the study.

Many researchers use lynx data to make performance comparisons between linear and nonlinear methods. Before modeling, the series' logarithm to the base 10 was computed, as noted by [40, 46]. The last 14 observations of the series were used to form a test set. The second series is the annual Real GNP of the USA, which exhibits a trend over time and has a strong cyclical pattern, as can be observed in the earlier period of the series [25]. The GNP series lacks a seasonal component and is non-stationary due to cyclic oscillations. The series has a total of 85 observations; we used 70 of them to model the series, and the remaining observations were used to form a test set. The third series consists of the annual number of births per 10,000 23-year-old women in the USA from 1917 to 1975.

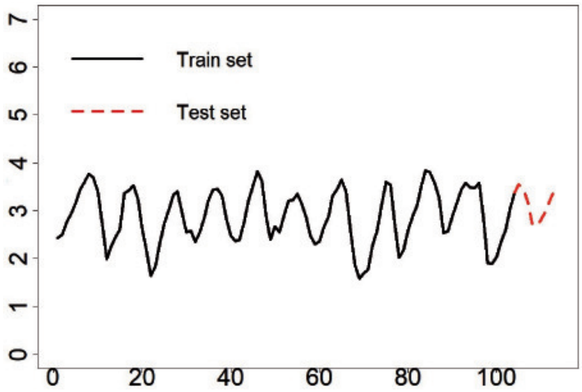



**Figure 1.** The plot of lynx

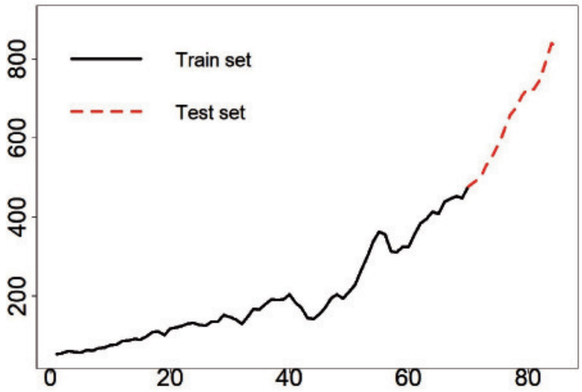



**Figure 2.** The plot of real gross national product

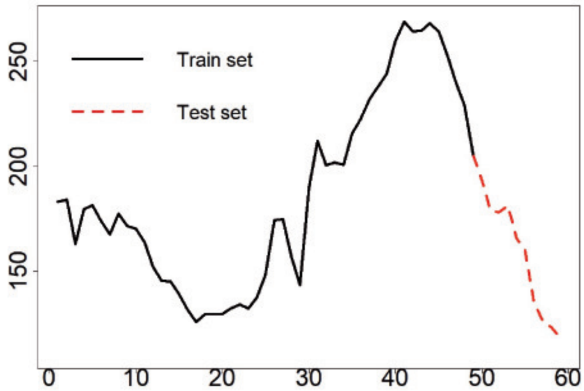



**Figure 3.** The plot of birth rate

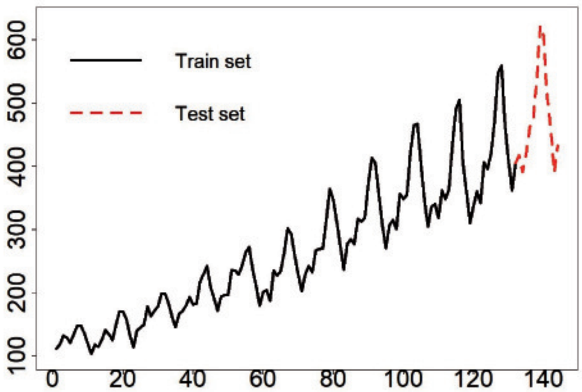



**Figure 4.** The plot of airline passengers

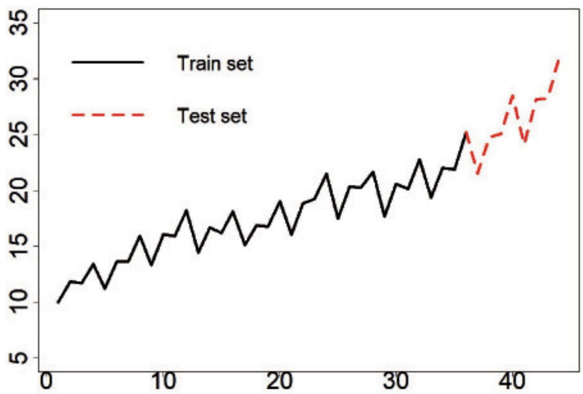



**Figure 5.** The plot of new plant or equipment Expend.s in USA

Figures 3–5 show the time plots of the series that are useful for choosing the most appropriate techniques in the field of forecasting. Table 2 gives the descriptive statistics of the time series.

The series' time plot vividly demonstrates the wide-ranging changes in the birthrate, which include a decline during the Great Depression, an increase following World War II, and a subsequent decline after 1960 [51]. The time plot reveals that this series is neither stationary nor seasonal. There are 59 observations in total. The test set includes the last ten observations of the series. The monthly passenger numbers in international air travel from 1949 to 1960 is another series that we have collected from the book by Box and Jenkins [10]. We took the number of passengers' logarithms in base 10 following the study by [10].

**Table 2.** Descriptive statistics

| Metrics | Lynx | RGNP | Birth rate | Passenger | Expend.s |
|---|---|---|---|---|---|
| Mean | 2.904 | 278.091 | 181.485 | 280.299 | 18.983 |
| Median | 2.887 | 189.800 | 174.700 | 265.500 | 18.515 |
| Standard deviation | 0.558 | 215.606 | 44.243 | 119.966 | 5.024 |
| Variance | 0.312 | 46486.002 | 1957.431 | 14391.917 | 25.243 |
| Kurtosis | –0.712 | 0.116 | –0.750 | –0.365 | –0.033 |
| Skewness | –0.367 | 1.079 | 0.553 | 0.583 | 0.503 |
| Range | 2.253 | 786.500 | 150.300 | 518.000 | 21.920 |
| Maximum | 1.591 | 52.700 | 118.500 | 104.000 | 10.000 |
| Minimum | 3.845 | 839.200 | 268.800 | 622.000 | 31.920 |
| Observation number | 114.000 | 85.000 | 59.000 | 144.000 | 44.000 |

The data is an example of a seasonal time series. Due to the seasonal nature of air travel, we expect more travel during the summer months, which is seen in the time plot of the series that shows a 12-month pattern and an upward trend [53]. The data consists of 144 observations representing the total number of international passengers per month. The most recent 12 observations are taken as the test set. The data on quarterly Expend.s in the USA for new plants and equipment from 1964 to 1976 is the final series examined in this study. The time plot of the series shows that the new plant or equipment expenses have a seasonal pattern. The original series is divided into two datasets: one with eight observations to make an out−of−sample comparison between forecasting models and the other with 44 observations to train the models. All datasets mentioned here can easily be found on the website of the Time Series Data Library[1].

## 3.2. Methods

In this section, we briefly describe the forecasting methods. In the subsequent section, the proposed algorithm is discussed in detail.

### 3.2.1. Forecasting models

We use five different single forecasting models for comparison purposes. These models include both linear and nonlinear models. The description of each model is presented below.

**Autoregressive integrated moving average (ARIMA) model.** As one of the most widely employed models, ARIMA was proposed by [10] and has found many application areas in forecasting time series by relying on the linear function of the lagged values and past errors. It is composed of three components: the autoregressive component, which takes into account the past observations of the variable to be forecasted, the integrated part to make the series stationary, and the moving average component to model lagged

[1] https://robjhyndman.com/tsdl/

forecast errors. The general form of the ARIMA model for non-seasonal time series can be defined as follows:

$$y_t^{'} = c + \theta_1 y_{t-1}^{'} + \cdots + \theta_p y_{t-p}^{'} - \phi_1 \varepsilon_{t-1} - \cdots - \phi_q \varepsilon_{t-q} + \varepsilon_t \quad (1)$$

where $y_t^{'}$ is the differenced variable, $\varepsilon_t$ represents the random error at time $t, c$ is the constant term, the coefficients $\theta_i\ (i = 1, ..., p)$ and $\phi_j\ (j = 1, ..., q)$ denotes autoregression coefficients and moving average coefficients, respectively.

The notation of ARIMA $p, d, q$ is frequently used to show the order of the autoregressive part, the required number of non-seasonal differencing operations, and the order of past errors. The ARIMA model depends on three stages of model building after checking the stationary assumption and making it stationary if it is not. The first one is the identification of the form of equation (1) by benefiting from autocorrelation and partial autocorrelation functions to determine the orders of$p$ and $q$. The second one is the estimation which includes parameter estimation of the selected models. The last one is diagnostic checking, which examines the residuals to ensure that they display white noise characteristics.

**Exponential smoothing (ETS) model.** Based on the fact that the impact of recent observations on the variable to be forecasted is higher than those of past observations, it is plausible to utilise exponential smoothing methods, which will assign exponentially more weight to the latest observations. Because it is simple to understand and easy to use, this family of methods has been used and has produced reliable forecasts in industry and economics for a long time [20]. Besides, these methods enable us to directly model the components of the level, trend, and seasonality. Depending on the way the components are dealt with, it is possible to classify exponential smoothing methods as in Table 2. Each combination corresponds to a specific exponential smoothing method. For example, $(A, N)$ stands for Holt's linear method, and $(A, M)$ represents the multiplicative Holt-Winters method.

**Table 3.** The ways of defining trend and seasonality

| Trend component | Seasonal component | | |
|---|---|---|---|
| | N (none) | A (additive) | M (multiplicative) |
| N (none) | N, N | N, A | N, M |
| A (additive) | A, N | A, A | A, M |
| $A_d$ (additive damped) | $A_d$, N | $A_d$, A | $A_d$, M |
| M (multiplicative) | M, N | M, A | M, M |
| $M_d$ (multiplicative damped) | $M_d$, N | $M_d$, A | $M_d$, M |

After the study by [38], which connects the formulations of state space to the methods of exponential smoothing, [29] presented likelihood calculations and model selection criteria for exponential smoothing methods. The acronym ETS was introduced to distinguish between different models. In this acronym, E corresponds to additive or multiplicative errors, T and S are, in turn, the trend and seasonal components. To select among 30 state space models, well-known information criteria like AIC, AICc, and BIC can be used. For more information about the ETS model, interested readers may refer to a book written by [28]

**Self-exciting threshold autoregressive (SETAR) model.** Time series can be composed of multiple regimes which describe different relationships, such as explosive or contractionary regimes, between independent and dependent variables. The first attempt to model multiple regimes goes back to a study by [6]. Later, [50] developed this idea by using the threshold autoregressive (TAR) model, which determines a regime by comparing an observable variable with a threshold value, and applying the linear autoregressive model to each regime. The self-exciting threshold autoregressive (SETAR) model is a particular case for the TAR model where regime switching depends on the self-dynamics of the dependent variable, for example, $y_{t-d}$ in which d is the length of the lag. A two−regime SETAR model can be represented as SETAR(2; $p$, $r$) and written algebraically as follows:

$$y_t = \begin{cases} \phi_0^{(1)} + \sum_{i=1}^{p} \phi_i^{(1)} y_{t-i} + \varepsilon_t^{(1)} & \text{if } y_{t-d} \leq \tau \\ \phi_0^{(2)} + \sum_{i=1}^{r} \phi_i^{(2)} y_{t-i} + \varepsilon_t^{(2)} & \text{if } y_{t-d} < \tau \end{cases} \tag{2}$$

where d≥0 denotes the delay parameter, $y_{t-d}$ represents the threshold variable, $\tau$ is the threshold value which divides the series into two parts, $\phi_i^{(1)}$ and $\phi_i^{(2)}$ are the coefficients of lower and upper regimes, and $p$ and $r$ show the order of the autoregressive model, $\varepsilon_t$ is assumed to follow $IID(0, \sigma^2)$.

A SETAR model with more regimes can be defined similarly. The parameters of autoregressive models can be estimated easily by the ordinary least square method. In addition, the value of the threshold and the parameters of the SETAR model must be identified, and a procedure consisting of three stages by [49] can be followed for this purpose.

**Logistic smooth transition autoregressive (LSTAR) model.** One way of handling regime-switching is the smooth transition autoregressive (STAR) model developed by [35, 47]. The key variation between SETAR and STAR models is in the process of controlling the transition between regimes. Instead of a discrete or discontinuous transition between regimes, as in the case of the TAR or SETAR models, the STAR model performs a smooth transition by using a transition function to reduce the speed of transition.

$$y_t = a_0 \sum_{i=1}^{p} a_i y_{t-i} + \left( \sum_{i=1}^{p} \beta_i y_{t-i} \right) F\left(z_{t-d}\right) + \varepsilon_t \tag{3}$$

where $z_{t-d}$ denotes the transition variable which can be determined endogenously or exogenously, $p$ is the order of the STAR model, and $F\left(z_{t-d}\right)$ corresponds to the transition function, which is smooth and continuous.

Two alternative transition functions are popular among researchers. One exploits equation (4) and is known as the logistic smooth transition autoregressive (LSTAR) model. The other one employs the exponential function given in equation (5) and is known as the exponential smooth transition autoregressive (ESTAR) model:

$$F\left(z_{t-d}\right) = \left[1 + e^{-\gamma(z_{t-d}-c)}\right]^{-1} \tag{4}$$

where $\gamma$ is the smoothing parameter governing the rate of transition between regimes, $c$ represents the threshold parameter, and $d$ is the delay parameter.

$$F\left(z_{t-d}\right) = \left[1 - e^{-\gamma(z_{t-d}-c)^2}\right], \gamma > 0 \tag{5}$$

The LSTAR model allows asymmetric realisations to characterise different dynamic behaviours in the regimes, while ESTAR only supports symmetric behaviour. Following the decision rule suggested by [48] to select between two models, we chose the LSTAR model in this study.

**Artificial neural networks (ANN) model.** Even though the first ANN model dates back to the 1940s, there has been a renewed surge of interest through the invention of the backpropagation algorithm by [43] since 1986. Thanks to this breakthrough algorithm, ANN gained the property of approximating any nonlinear functions at the desired level provided a sufficient number of neurons is added to the hidden layer [15]. ANN does not require any assumption about data characteristics. Among a wide variety of ANN types, multi-layer perceptron (MLP) has stood out from the others and has successfully modelled time series in numerous papers [56]. The general functional form of the ANN model is defined in Eq. 6, as follows:

$$y_t = w_0 + \sum_{j=1}^{q} w_j f\left(w_{0j} + \sum_{i=1}^{p} w_{ij} y_{t-i}\right) + \varepsilon_t \tag{6}$$

where $w_0$ and $w_{0j}(j = 1, \ldots, q)$ denote the biases, $w_j$ and $w_{ij}(i = 1, \ldots, p; j = 1, \ldots, q)$ represent the connection weights of input and hidden layers, $f$ is the transfer function, $p$ and $q$ are the numbers of neurons in the input and hidden layers, respectively.

The major drawback of ANN is the problem of overfitting to peculiar properties of data or memorising the noise that existed in the training set. In the autoregressive ANN model, the effect of the numbers of input and hidden neurons on the out−of−sample performance is crucial [4].

### 3.2.2. Combining algorithms

First, we briefly introduce the most common and simple combining algorithms, and then we discuss some issues that are widely encountered in practice.

**Mean combining.** The most sensible approach to combine forecasts is to take the average of all those forecasts. This approach is a superior benchmark because of its simplicity and utility [21]. It should be noted that the Mean combining algorithm shows higher forecasting accuracy when the forecasting performances of the models are slightly different:

$$f^c = \frac{1}{n} \sum_{i=1}^{n} f_i \tag{7}$$

where $f^c$ denotes the forecast of the combining method, $f_i$ represents the forecast of the $i^{th}$ forecasting model, and n is the number of single models.

**Median combining.** Another simple and intriguing combining algorithm is to use the median of all forecasts. However, Median combining is insensitive to the variation in model performances. Therefore, if one of the models has a superior forecasting performance compared to the others, this algorithm may not be a suitable combining method [39]. The forecast equation for Median combining is as follows:

$$f^c = \begin{cases} f_{(\frac{n}{2}+0.5)} & \text{if } n \text{ is odd} \\ \frac{1}{2}\left(f_{(\frac{n}{2})} + f_{(\frac{n}{2}+1)}\right) & \text{if } n \text{ is even} \end{cases} \tag{8}$$

**Trimmed mean combining**. The trimmed mean combining approach is a robust method for the outliers as it excludes the smallest and the largest forecasts before computing the arithmetic average of all individual forecasts. The final combined forecast is calculated by using a trim factor $\lambda$ as follows:

$$f^c = \frac{1}{n(1-2\lambda)} \sum_{i=\lambda n+1}^{(1-\lambda)n} f_i \tag{9}$$

**procedure** INVERSE RANK WEIGHTING SCHEME
    $y_i \leftarrow$ *Input*
    $i \leftarrow$ *the number of datasets*
$\hat{f}$:
    **for** *every search i*
*calculate forecasts* $\hat{f}_j$ *in* $\Omega_j$
    $j \leftarrow$ *the number of forecasting models*
    $\Omega_j \leftarrow$ *the set of forecasting models*
*MSE*:
    **for** *every search j*
*obtain* $\hat{\varepsilon}_{nj}$ *based on the difference between* $y_{ni}$ *and* $\hat{f}_{nj}$
*calculate* $MSE_j$ *using* $\hat{\varepsilon}_{nj}$
    $n \leftarrow$ *the number of observations in test set*
*R*:
    **if** $MSE_j$ has the smallest value **then return** $R_j$ gets rank 1, and the following smallest $R_j$ gets rank 2, etc.
    **end if**
    **Save** $R_j$
$\omega$:
*calculate weights* $\omega_j$ *using* $R_j$
    $\omega_j = \dfrac{R_j^{-1}}{\sum_{i=1}^{j} R_j^{-1}}$
**end procedure**

**Algorithm 1.** Inverse rank weighting scheme

**Equally weighted algorithm betwee n mean and median combining.** Another robust weighting scheme is to use the advantages of mean and median combining methods by utilising the equal weights rule that offers reasonably improved forecast accuracy

$$f^c = \alpha_i f_{c,\,\text{mean}} + (1-\alpha_i)\, f_{c,\text{median}} \tag{10}$$

In this algorithm, when $\alpha = 0.5$, it provides equal weights for the combination between the mean and median combining [1]. From this viewpoint, it performs a balancing task between the two methods.

**Inverse rank.** The combining algorithm by using the inverse ranks was proposed by [2]. A robust weighting scheme, which is expected to rank mean square errors (MSE) in the reverse order, is determined as the proportion of each reversed rank order by the sum of reversed orders. The pseudo−code for this method is presented in Algorithm 1.

**A weighting scheme based on direct forecast.** The weighting scheme in this algorithm is based on the accuracy of the sign of the forecast direction [5]. As is known, the best forecasts do not only mean attaining the minimum forecast error measured by MSE but also producing movements that are in harmony with the real values of a time series. Here, we modified the algorithm proposed by [5] regarding the calculation of the binary matrix. The second binary matrix denoted by $S_{tj}^{(2)}$ is added to the analysis to see that the movements calculated by the difference between two consecutive forecasts may be a better measure of forecast accuracy. The details of this method for obtaining the weights of each forecasting model are provided in Algorithm ??.

**The proposed algorithm**. The proposed algorithm is composed of two procedures, namely, inverse ranks and forecast direction. In the first procedure, we calculated the MSE metric for each forecasting model ranked the models according to their performance, and used inverse ranks to obtain the weights of the models in the combining method. In the second procedure, if the sign function of the difference between $Y_{t-1}$ and $Y_{t-2}$ is equal to the sign function of the difference between the estimated $\hat{Y}_{t-1}$ and $Y_{t-2}$, then it takes the value zero; otherwise, it takes the value one.

**procedure** A ROBUST WEIGHTING SCHEME BY FORECAST DIRECTION
    $y_i \leftarrow$ *Input*
    $i \leftarrow$ *the number of datasets*
$\hat{y}$:
    **for** *every search i*
*calculate predictions* $\hat{y}_n$ *in* $\Omega_j$
    $j \leftarrow j^{th}$ *forecasting model*
    $n \leftarrow$ *the number of observations in training set*
    $\Omega_j \leftarrow$ *the set of forecasting models*
$S$:
    **for** *two binary matricies*

$$S_{tj}^{(1)} = \begin{cases} if \quad sign\left(y_{(t-1),j} - y_{(t-2),j}\right) = sign\left(\hat{y}_{(t-1),j} - y_{(t-2),j}\right), 0 \\ if \quad sign\left(y_{(t-1),j} - y_{(t-2),j}\right) \neq sign\left(\hat{y}_{(t-1),j} - y_{(t-2),j}\right), 1 \end{cases}$$

$$S_{tj}^{(2)} = \begin{cases} if \quad sign\left(y_{(t-1),j} - y_{(t-2),j}\right) = sign\left(\hat{y}_{(t-1),j} - \hat{y}_{(t-2),j}\right), 0 \\ if \quad sign\left(y_{(t-1),j} - y_{(t-2),j}\right) \neq sign\left(\hat{y}_{(t-1),j} - \hat{y}_{(t-2),j}\right), 1 \end{cases}$$

    $t \leftarrow 1, \ldots, n$
    $FP \leftarrow$ *the forecasting performance in training set*
    **while** FP($S_{tj}^{(1)}$) is better than FP($S_{tj}^{(2)}$) **do** select $S_{tj}^{(1)}$
    **end while**1
    **while** FP($S_{tj}^{(2)}$) is better than FP($S_{tj}^{(1)}$) **do** select $S_{tj}^{(2)}$
    **end while**1
$\omega$:
    *calculate weights* $\omega_{tj}$ *using* $S_{tj}$

$$\omega_{tj} = \frac{exp\left(-sum_{i=1}^{t-1} S_{ij}\right)}{\sum_{j=1}^{\Omega_j} \exp\left(-sum_{i=1}^{t-1} S_{ij}\right)}$$

**end procedure**

**Algorithm 2.** A robust weighting scheme by forecast direction

Moreover, for the second binary matrix in the study, if the sign function of the difference between $Y_{t-1}$ and $Y_{t-2}$ is equal to the sign function of the difference between the estimated $\hat{Y}_{t-1}$ and the estimated $\hat{Y}_{t-2}$, then it takes the value zero; otherwise, it takes the value one. We created a dummy variable that included zeros and ones, and if the number of ones was less than the number of zeros in the model, we gave a rank of one. As a final step, we combined these two ranks to obtain the weights that take into account the benefits of two combining methods as a meta−combiner. The proposed algorithm is defined in Algorithm 2.

### 3.2.3. Loss functions

A loss function in forecasting studies measures how well a given forecasting model fits the real data. In the literature, it can be seen that the effects of using various accuracy metrics may vary in terms of selected error metrics [44]. Accordingly, the selection of appropriate error metrics is one of the performance evaluation parameters. To show the ability of the proposed algorithm in the five given different datasets, the most widely used metrics are calculated and their definitions are given below.

**Mean square error (MSE).** The simplest and most frequently used loss function in the literature is the MSE. The MSE is defined by the following equation:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}\left(f_t - \hat{f}_t\right) \tag{11}$$

**Mean absolute error (MAE).** Another one of the existing error measures in the literature is MAE, which is also used as another loss function to compare the out-of-sample forecasting performances of the models. A large body of literature uses the popular MAE loss function to avoid the effects of outliers in forecasts.

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|f_t - \hat{f}_t\right| \tag{12}$$

**Geometric mean square error (gMSE).** Another robust loss function to avoid outliers is the gMSE. The main advantage of using the gMSE loss function is that the mean absolute errors of different models can be compared by computing their geometric means.

$$gMSE = \sqrt[n]{\prod_{t=1}^{n}\left(f_t - \hat{f}_t\right)} \tag{13}$$

**Mean absolute percentage error (MAPE).** It is a good option for studying data that contains extreme values since it is less sensitive to outliers than other error measures like the MAE or square root of the MSE.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{f_t - \hat{f}_t}{f_t}\right| \tag{14}$$

**procedure** A NEW WEIGHTING SCHEME BY A META-COMBINING METHOD
  $s \leftarrow$ *the number of binary matrices*
  $t \leftarrow 1, \ldots, n$
  $j \leftarrow j^{th}$ *forecasting model*
  $\nu \leftarrow$ *the number of forecasting models*
  *Input* $\leftarrow R_{k1}$ *and* $R_{k2}$
   $R_{k1} \leftarrow$ *the ranks from Algorithm I*
   $R_{k2} \leftarrow$ *the ranks from Algorithm II*
  $A_j \leftarrow$ *the total number of the wrong direction*
  **while** $S_{tj}^{s}$ **do**
   calculate $A_j = sum_{i=1}^{t-1} S_{ij}$ **for** each j
   **if** $A_j < A_w, j, w \in \Omega$ ; where $j \neq w$
   $R_{j,k2} \leftarrow$ *gets rank 1, if the number of 1's is less than the number of 0's in* $\Omega$
    **repeat**
   assign ranks to $R_{j,k2}$
    **until** $j = 1, \ldots, \nu$
  **end while**
  $\omega_{tj}$:
   calculate weights $\omega_{tj}$ using $R_{j,k1}$ and $R_{j,k2}$

$$\omega_{tj} = \frac{(R_{j,k1} + R_{j,k2})^{-1}}{\sum_{j=1}^{\nu} (R_{j,k1} + R_{j,k2})^{-1}}$$

  **Save** $\omega_{tj}$
**end procedure**

**Algorithm 3.** The proposed algorithm

# 4. Results and discussions

The methodology consists of guidance on modelling the data sets using univariate time series models, the ANN model and the proposed algorithm. The aim of the study is to identify the most accurate time series forecasting model among several forecasting models and to increase the forecasting accuracy by using the proposed the proposed hybrid algorithm. For this purpose, the forecasting results of the study are provided in Tables 4–7. The rows contain the individual models and the combining procedures, while the columns display the error measures obtained for each data set and their ranks among all forecasting methods. The last column represents the average rank computed by ranks for all data sets. Bold numbers in these tables indicate the best-performing model.

Three different error measures were employed to evaluate model performances. Each model was assessed within each error measure, and rank ordering was established accordingly. However, a direct comparison of the error measures or their superiority was not conducted. As seen in various literature studies [26, 32], different error measures can highlight the performance of different models. Similar findings emerged in this study as well.

Nevertheless, when the models' performances were ranked within their respective error measures, it became easier to discern which models stood out under different error measures. By taking the average of these ranks, comparing model performances becomes more informative, indicating which model presents less risk in terms of forecasting performance compared to others. In essence, the averaging of ranks provides readers with insights into the relative performance of models when assessed under different error measures on datasets

**Table 4.** Out-of-sample forecasting performances in terms of MSE

| Models and algorithms | Lynx | $R_1$ | RGNP | $R_2$ | Birth rate | $R_3$ | Passenger | $R_4$ | Expend. | $R_5$ | $\bar{R}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ARIMA | 0.0300 | 11 | 526.4260 | 12 | 136.5890 | 12 | 0.00175 | 5 | 0.6174 | 10 | 10 |
| SETAR | 0.0276 | 10 | 516.0593 | 11 | 96.1400 | 8 | 0.00210 | 11 | 0.5740 | 8 | 9.6 |
| LSTAR | 0.0268 | 9 | 426.1781 | 9 | 99.6308 | 10 | **0.00150** | **1** | 2.0475 | 12 | 8.2 |
| ANN | 0.0118 | 5 | 422.1891 | 8 | 115.6045 | 11 | 0.00177 | 6 | 0.5936 | 9 | 7.8 |
| ETS | 0.0477 | 12 | **376.0186** | **1** | 96.9108 | 9 | 0.00569 | 12 | 1.0485 | 11 | 9 |
| Mean combining | 0.0116 | 4 | 402.0424 | 5 | 83.4850 | 4 | 0.00199 | 10 | 0.4655 | 6 | 5.8 |
| Median combining | **0.0100** | **1** | 441.6993 | 10 | 82.4695 | 2 | 0.00195 | 8 | 0.4342 | 2 | 4.6 |
| Trimmed combining | 0.0119 | 6 | 417.0290 | 6 | **82.1847** | **1** | 0.00188 | 7 | 0.4345 | 3 | 4.6 |
| Equal weighted between mean and median | 0.0104 | 2 | 420.2410 | 7 | 82.4798 | 3 | 0.00196 | 9 | **0.4105** | **1** | 4.4 |
| Weighting algorithm [2] | 0.0147 | 7 | 397.3390 | 4 | 89.6471 | 6 | 0.00172 | 2 | 0.4513 | 5 | 4.8 |
| Weighting algorithm [5] | 0.0243 | 8 | 385.6306 | 2 | 94.8965 | 7 | 0.00174 | 4 | 0.5569 | 7 | 5.6 |
| Proposed algorithm | 0.0108 | 3 | 385.7200 | 3 | 86.6891 | 5 | 0.00173 | 3 | 0.4353 | 4 | **3.6** |

**Table 5.** Out-of-sample forecasting performances in terms of MAE

| Models and algorithms | Lynx | $R_1$ | RGNP | $R_2$ | Birth rate | $R_3$ | Passenger | $R_4$ | Expend. | $R_5$ | $\bar{R}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ARIMA | 0.133 | 10 | 17.287 | 10 | 10.150 | 12 | 0.03113 | 4.5 | 0.670 | 10 | 9.3 |
| SETAR | 0.137 | 11 | 18.402 | 12 | 7.511 | 10 | 0.03430 | 10 | 0.564 | 7 | 10 |
| LSTAR | 0.115 | 8 | 17.063 | 9 | 8.103 | 11 | **0.02780** | **1** | 1.247 | 12 | 8.2 |
| ANN | 0.093 | 6 | 17.317 | 11 | 6.941 | 7 | 0.03240 | 7 | 0.604 | 8 | 7.8 |
| ETS | 0.168 | 12 | **15.212** | **1** | 7.364 | 9 | 0.05670 | 12 | 0.852 | 11 | 9 |
| Mean combining | 0.090 | 5 | 16.217 | 5 | 6.093 | 3 | 0.03450 | 11 | 0.539 | 6 | 6 |
| Median combining | **0.083** | **1** | 17.041 | 8 | 6.393 | 6 | 0.03320 | 6 | **0.472** | **1** | 4.4 |
| Trimmed combining | 0.087 | 4 | 16.594 | 6 | 6.319 | 5 | 0.03250 | 8 | 0.496 | 3 | 5.2 |
| Equal weighted between mean and median | 0.085 | 3 | 16.629 | 7 | 6.159 | 4 | 0.03350 | 9 | 0.475 | 2 | 5 |
| Weighting algorithm [2] | 0.097 | 7 | 16.203 | 4 | **5.870** | **1** | 0.03111 | 2 | 0.511 | 5 | 3.8 |
| Weighting algorithm [5] | 0.119 | 9 | 15.895 | 2 | 7.310 | 8 | 0.03113 | 4.5 | 0.641 | 9 | 6.5 |
| The proposed algorithm | 0.084 | 2 | 15.900 | 3 | 5.910 | 2 | 0.03112 | 3 | 0.501 | 4 | **2.8** |

**Table 6.** Out-of-sample forecasting performances in terms of gMSE

| Models and algorithms | Lynx | $R_1$ | RGNP | $R_2$ | Birth rate | $R_3$ | Passenger | $R_4$ | Expend. | $R_5$ | $\bar{R}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ARIMA | 0.00620 | 10 | 119.563 | 10 | 63.391 | 12 | 0.00054 | 10 | 0.301 | 10 | 10.4 |
| SETAR | 0.00780 | 12 | 124.001 | 11 | 24.780 | 9 | 0.00034 | 2 | 0.063 | 2 | 7.2 |
| LSTAR | 0.00510 | 8 | 83.925 | 2 | 35.961 | 11 | **0.00025** | **1** | 0.988 | 12 | 6.8 |
| ANN | 0.00580 | 9 | 131.377 | 12 | 6.374 | 4 | 0.00052 | 7 | 0.150 | 8 | 8 |
| ETS | 0.00630 | 11 | **82.782** | **1** | 16.249 | 8 | 0.00134 | 12 | 0.370 | 11 | 8.6 |
| Mean combining | 0.00430 | 7 | 106.336 | 8 | 6.216 | 3 | 0.00067 | 11 | 0.131 | 7 | 7.2 |
| Median combining | 0.00363 | 2 | 104.329 | 7 | 11.652 | 6 | 0.00050 | 6 | **0.044** | **1** | 4.4 |
| Trimmed combining | **0.00304** | **1** | 96.156 | 6 | 12.250 | 7 | 0.00044 | 5 | 0.098 | 6 | 5 |
| Equal weighted between mean and median | 0.00377 | 6 | 115.146 | 9 | 8.868 | 5 | 0.00053 | 8.5 | 0.078 | 3 | 6.3 |
| Weighting algorithm [2] | 0.00372 | 5 | 95.226 | 5 | **2.510** | **1** | 0.00037 | 3 | 0.095 | 5 | 3.8 |
| Weighting algorithm [5] | 0.00370 | 4 | 84.703 | 3 | 25.140 | 10 | 0.00053 | 8.5 | 0.293 | 9 | 6.9 |
| The proposed algorithm | 0.00364 | 3 | 93.307 | 4 | 4.425 | 2 | 0.00039 | 4 | 0.082 | 4 | **3.4** |

exhibiting distinct characteristics. Consequently, it offers valuable information for model selection and decision-making by providing an understanding of how models perform in terms of forecasting under different scenarios.

**Table 7.** Out-of-sample forecasting performances in terms of MAPE

| Models and algorithms | Lynx | $R_1$ | RGNP | $R_2$ | Birth rate | $R_3$ | Passenger | $R_4$ | Expend. | $R_5$ | $\bar{R}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ARIMA | 0.0440 | 10 | 1.2184 | 12 | 2.5179 | 12 | 0.0050 | 2.5 | 0.0236 | 10 | 9.3 |
| SETAR | 0.0443 | 11 | 1.1371 | 11 | **0.8904** | **1** | 0.0056 | 10.5 | 0.0197 | 7 | 8.1 |
| LSTAR | 0.0381 | 8 | **0.8498** | **1** | 1.7989 | 10 | **0.0046** | **1** | 0.0426 | 12 | 6.4 |
| ANN | 0.0322 | 6 | 0.8773 | 2 | 1.2889 | 5 | 0.0053 | 6.5 | 0.0214 | 8 | 5.5 |
| ETS | 0.0577 | 12 | 0.9813 | 8 | 2.2048 | 11 | 0.0092 | 12 | 0.0305 | 11 | 10.8 |
| Mean combining | 0.0304 | 5 | 0.9630 | 6 | 1.4329 | 6 | 0.0056 | 10.5 | 0.0194 | 6 | 6.7 |
| Median combining | 0.0283 | 2 | 1.0296 | 10 | 1.4927 | 8 | 0.0054 | 8 | **0.0169** | **1** | 5.8 |
| Trimmed combining | 0.0296 | 4 | 0.9727 | 7 | 1.5693 | 9 | 0.0053 | 6.5 | 0.0181 | 3.5 | 6 |
| Equal weighted between mean and median | 0.0289 | 3 | 0.9963 | 9 | 1.4494 | 7 | 0.0055 | 9 | 0.0172 | 2 | 6 |
| Weighting algorithm [2] | 0.0326 | 7 | 0.9108 | 4 | 1.1734 | 3 | 0.0051 | 4.5 | 0.0185 | 5 | 4.7 |
| Weighting algorithm [5] | 0.0394 | 9 | 0.9095 | 3 | 0.9258 | 2 | 0.0051 | 4.5 | 0.0226 | 9 | 5.5 |
| The proposed algorithm | **0.0275** | **1** | 0.9426 | 5 | 1.1773 | 4 | 0.0050 | 2.5 | 0.0181 | 3.5 | **3.2** |

As can be seen in Tables 4–7, the proposed algorithm has the minimum average rank value for the three different loss functions that have been considered. Therefore, choosing the proposed algorithm instead of the individual models and other combining algorithms provides more accurate forecasts in terms of ranks compared with these methods. Even if the proposed combining method does not lead to the best forecasting performance for all data sets and error measures, the overall performance, denoted by $\overline{R}$, is the best one as it demonstrates reliability and generality. To ensure a fair comparison of the model performances obtained in the study, the median value of ranks instead of the rank averages for the models has also been calculated and provided in Tables 8–11. Based on the MSE error metric that we proposed in the study, the combining algorithm places the models' rank performances as the second best according to the median value. However, when considering the error metrics MAE, MAPE and gMSE, it is observed that it has the lowest rank median value.

**Table 8.** Out-of-sample forecasting performances
in terms of MSE with the median ranks

| Models and algorithms | Medians of ranks |
|---|---|
| ARIMA | 11 |
| SETAR | 10 |
| LSTAR | 9 |
| ANN | 8 |
| ETS | 11 |
| Mean combining | 5 |
| **Median combining** | **2** |
| Trimmed combining | 6 |
| Equal weighted between mean and median | 3 |
| Weighting algorithm proposed by [2] | 5 |
| Weighting algorithm proposed by [5] | 7 |
| Proposed algorithm | 3 |

In this context, it is observed that selecting the proposed combining algorithm for out-of-sample forecasting performance across different datasets tends to reduce the risk of achieving lower forecasting accuracy compared to other models. As shown in Figures 6–9, the proposed algorithm reduces the risk of misjudgment in decision-making by providing the minimum loss functions among most of the datasets.

**Table 9.** Out-of-sample forecasting performances in terms of MAE with the median ranks

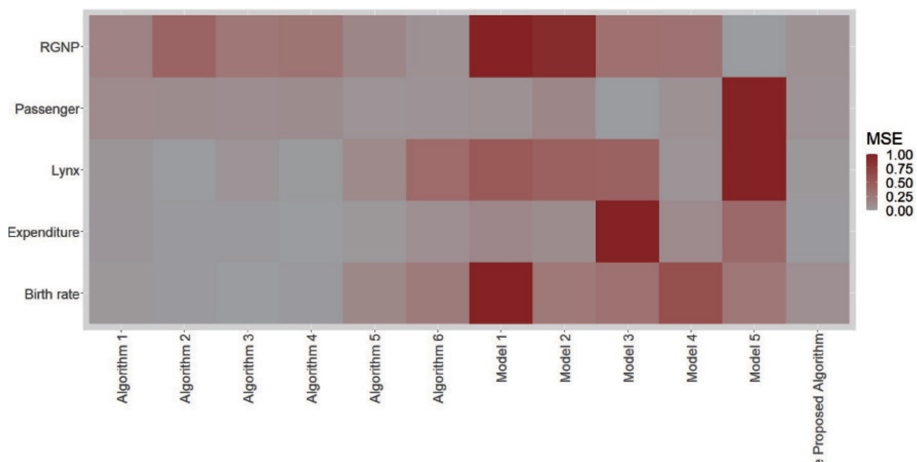
| Models and algorithms | Medians of ranks |
|---|---|
| ARIMA | 10 |
| SETAR | 10 |
| LSTAR | 9 |
| ANN | 7 |
| ETS | 11 |
| Mean combining | 5 |
| Median combining | 6 |
| Trimmed combining | 5 |
| Equal weighted between mean and median | 4 |
| Weighting algorithm proposed by [2] | 4 |
| Weighting algorithm proposed by [5] | 8 |
| **Proposed algorithm** | **3** |

**Table 10.** Out-of-sample forecasting performances in terms of gMSE with the median ranks

| Models and algorithms | Medians of ranks |
|---|---|
| ARIMA | 10 |
| SETAR | 9 |
| LSTAR | 8 |
| ANN | 8 |
| ETS | 11 |
| Mean combining | 7 |
| Median combining | 6 |
| Trimmed combining | 6 |
| Equal weighted between mean and median | 6 |
| Weighting algorithm proposed by [2] | 5 |
| Weighting algorithm proposed by [5] | 8.5 |
| **Proposed algorithm** | **4** |

**Table 11.** Out-of-sample forecasting performances in terms of MAPE with the median ranks

| Models and algorithms | Medians of ranks |
|---|---|
| ARIMA | 10 |
| SETAR | 10.5 |
| LSTAR | 8 |
| ANN | 6 |
| ETS | 11 |
| Mean combining | 6 |
| Median combining | 8 |
| Trimmed combining | 6.5 |
| Equal weighted between mean and median | 7 |
| Weighting algorithm proposed by [2] | 4.5 |
| Weighting algorithm proposed by [5] | 4.5 |
| **Proposed algorithm** | **3.5** |



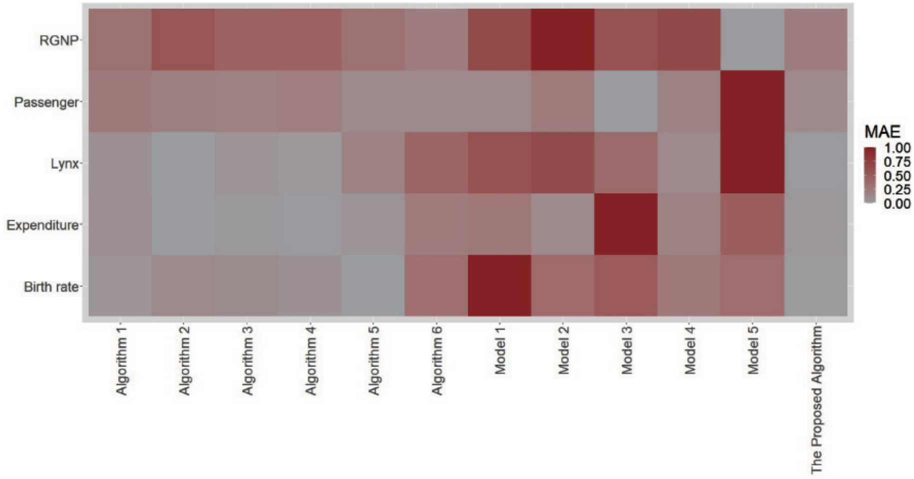



**Figure 7.** Heatmap of forecasting methods in terms of MAE

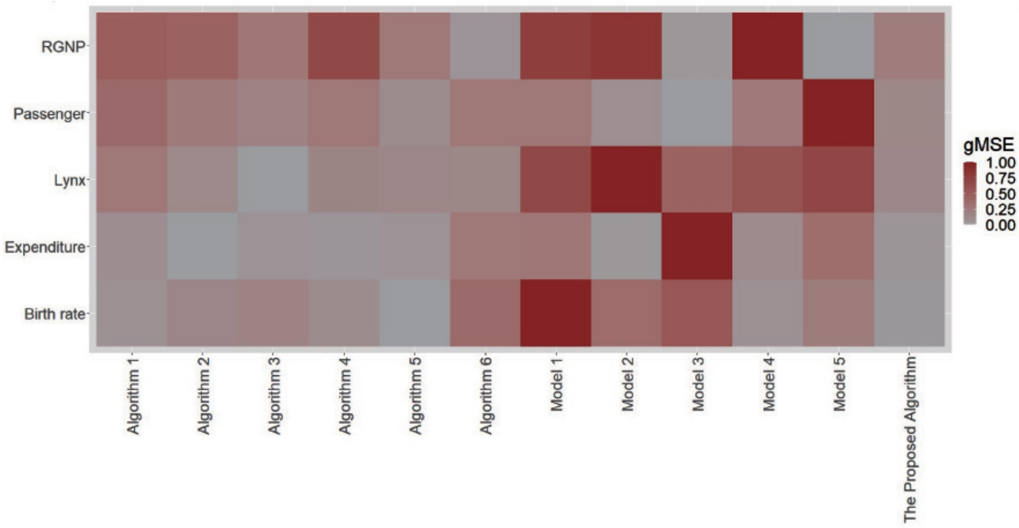



**Figure 8.** Heatmap of forecasting methods in terms of gMSE

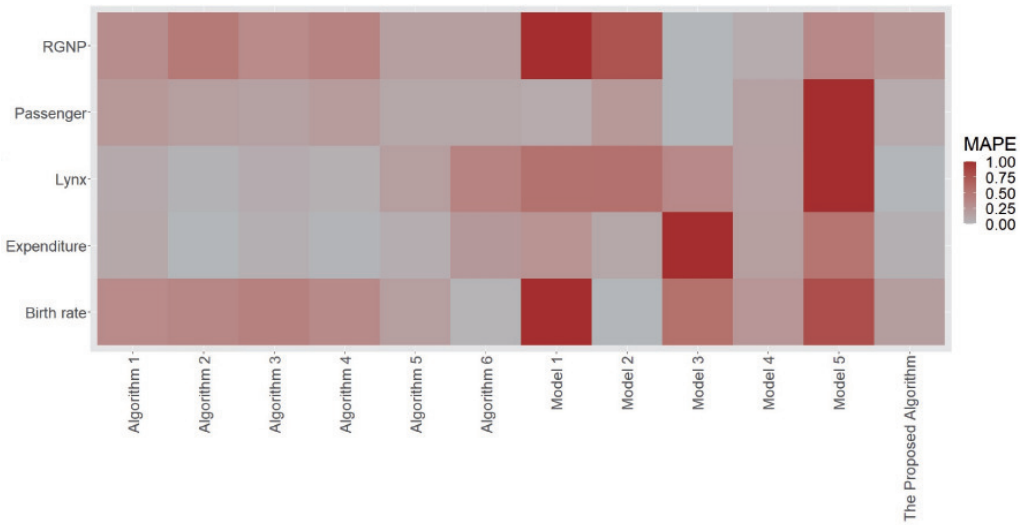



**Figure 9.** Heatmap of forecasting methods in terms of MAPE

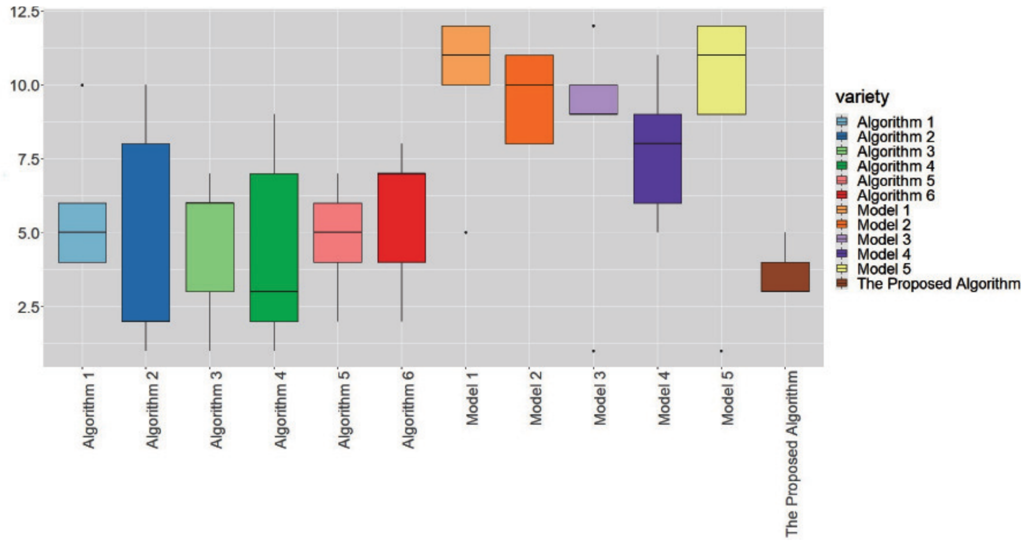



**Figure 10.** The Boxplot of the ranks in terms of MSE

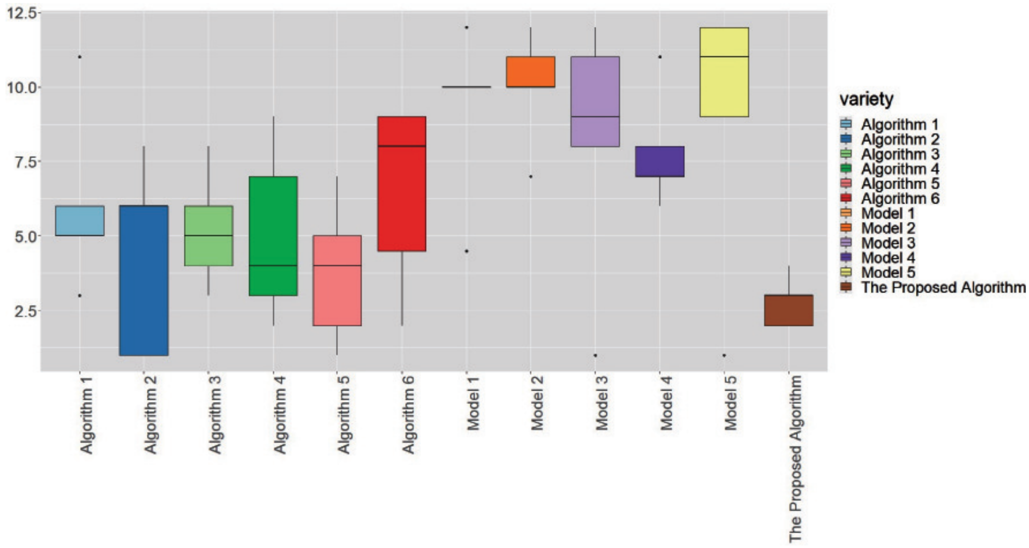



**Figure 11.** The Boxplot of the ranks in terms of MAE

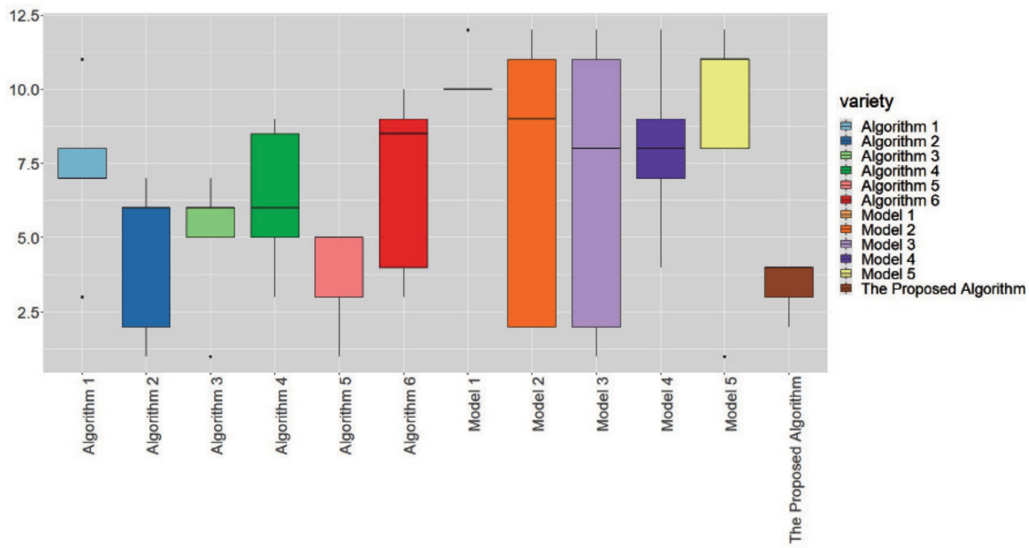



**Figure 12.** The Boxplot of the ranks in terms of gMSE

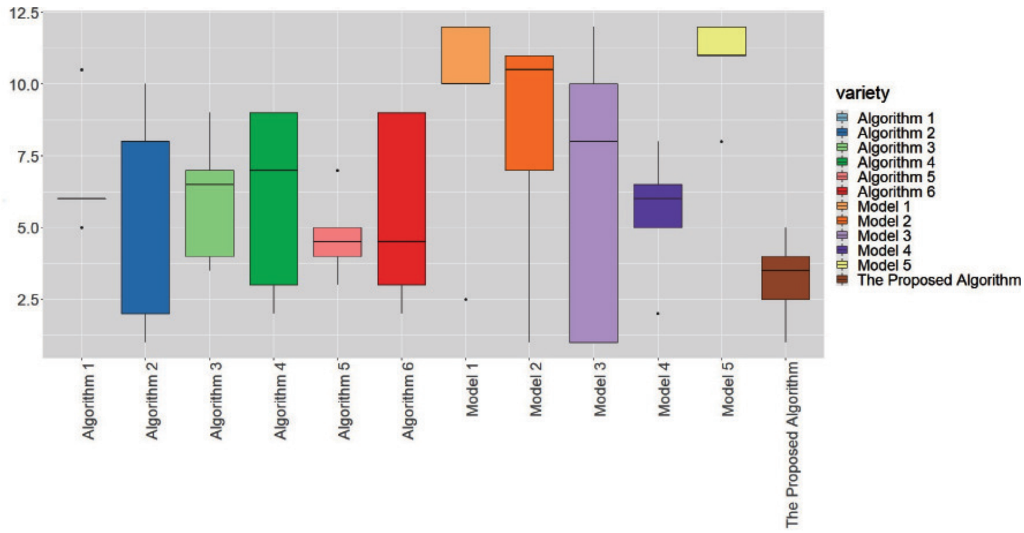



**Figure 13.** The Boxplot of the ranks in terms of MAPE

The abbreviations of the forecasted models and the algorithms stand for model 1 – ARIMA, model 2 – SETAR, model 3 – LSTAR, model 4 – ANN, model 5 – ETS, algorithm 1 – mean combining, algorithm 2 – median combining, algorithm 3 – trimmed combining, algorithm 4 – equal weighted combining, algorithm 5 – weighting based on the direct forecast, and algorithm 6 – weighting based on inverse rank.

In Figures 10–13, the boxplots of the ranks better display comparisons between the methods by using three loss functions. The minimum average rank of the forecasting models and the combining algorithms was achieved by using the proposed combining algorithm. Increasing the forecast accuracy in the proposed algorithm helps to obtain competitive and outperforming forecasts. As a result, the proposed algorithm needed less time in the decision-making process, had enhanced stability, and achieved high forecast accuracy by providing promising results for all datasets.

The result of the Friedman test in Figures 14–17 shows that the mean scores of ranks in representative forecasting models and combining algorithms are not equal in terms of different loss functions. The proposed algorithm achieves the best rank in terms of the Friedman test.

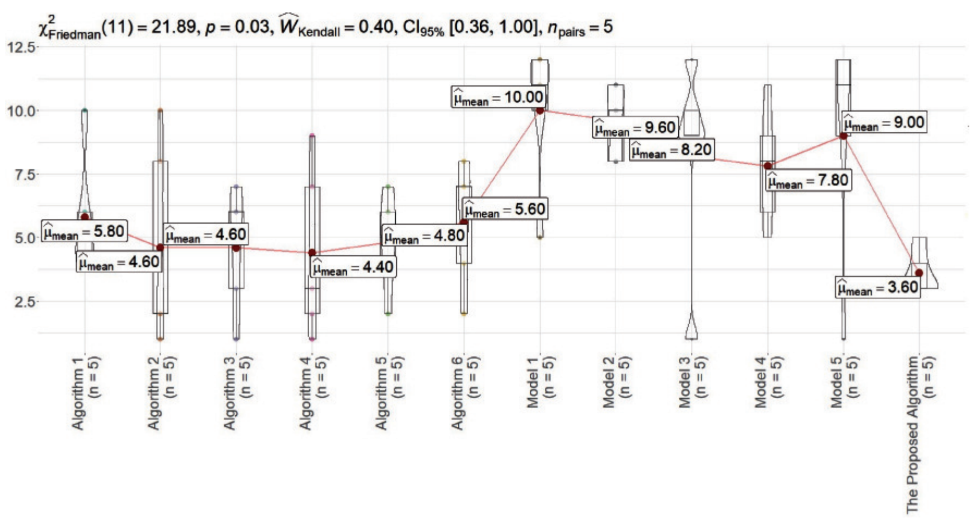



**Figure 14.** The Friedman test in terms of comparisons of the ranks by MSE

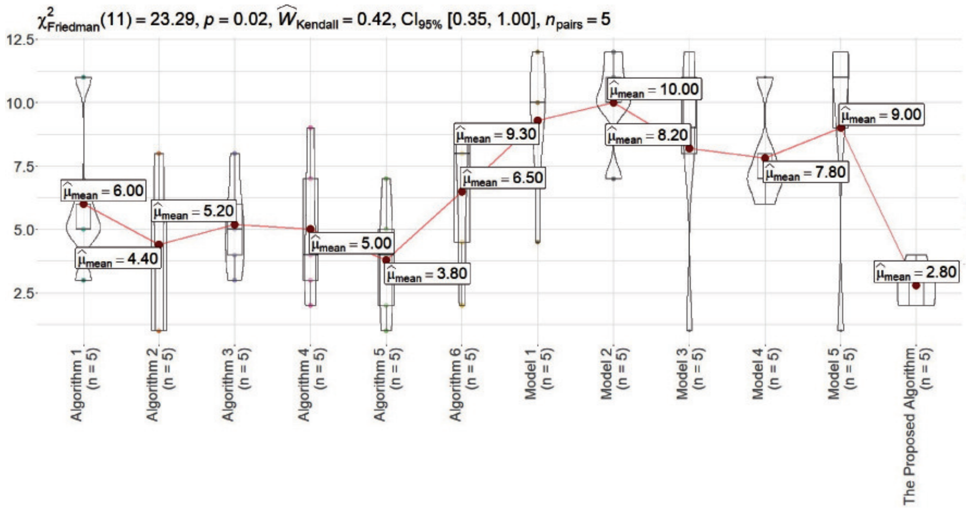



**Figure 15.** The Friedman test in terms of comparisons of the ranks by MAE

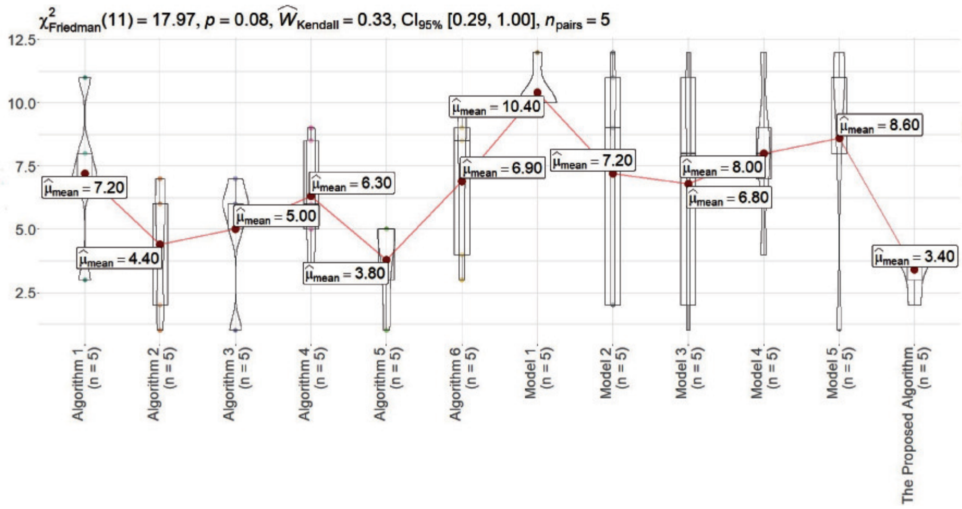



**Figure 16.** The Friedman test in terms of comparisons of the ranks by gMSE

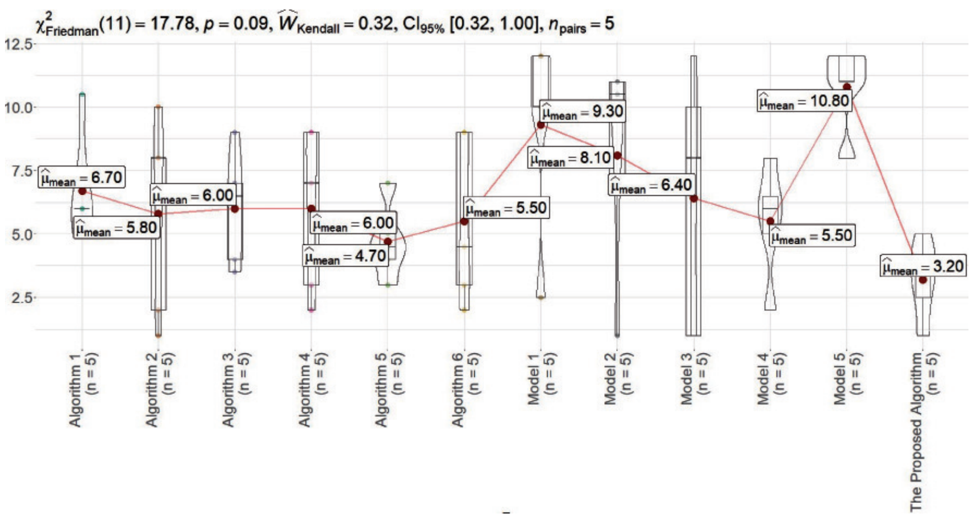



**Figure 17.** The Friedman test in terms of comparisons of the ranks by MAPE

The findings demonstrate that there is a difference between the forecasting performances of the proposed algorithm and the compared combining algorithms with a Kendall's $W$ score of around 0.4, which confirms the moderate agreement in five datasets.

# 5. Conclusions

This study introduces a meta-combining algorithm that performs well in all data sets by determining the weights and considering the information in the minimum squared error and the maximum forecast direction at the same time. Because it relies on the ranks calculated from two criteria, this algorithm can be regarded as a robust weighting scheme. By creating algorithms that attempt to combine for improved accuracy, the algorithm adopts the improvements to enhance forecasting performance. The forecast improvements arise from two sources. First, applying the weighting algorithm considered the best-fit model to time series data effectively reduces forecast errors. Second, using a weighting combining algorithm based on forecast direction performance is essential for the success of forecast combinations. In addition, the proposed algorithm incorporates the advantage of combining algorithm diversity with different algorithms capturing the effectiveness of the forecasting procedure.

In doing this, we show the applicability of a meta-combining approach in time series forecasting problems. This study also indicates that the forecast−combining algorithms can be compressed into a single-algorithm approach to yield more accurate results. This new combining algorithm−based on weighting retains a competitive predictive performance when compared with the other forecasting models and combining algorithms while being simpler and producing more reliable decisions.

For future directions, a more complicated meta−combining method, which requires parameter estimation to determine the weights in the combining method, may be developed. In addition, the form of the combining equation can be changed to generate combined forecasts with a nonlinear structure. Lastly, more data sets from different fields may be used to prove the value of the proposed method.

# References

[1] Adhikari, R., and Agrawal, R. K. A linear hybrid methodology for improving accuracy of time series forecasting. *Neural Computing and Applications 25*, 2 (2014), 269–281.

[2] Aiolfi, M., and Timmermann, A. Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics 135*, 1-2 (2006), 31–53.

[3] Aras, S., and Gulay, E. A new consensus between the mean and median combination methods to improve forecasting accuracy. *Serbian Journal of Management 12*, 2 (2017), 217–236.

[4] Aras, S., and Kocakoç, İ. D. A new model selection strategy in time series forecasting with artificial neural networks: IHTS. *Neurocomputing 174*, Part B, (2016), 974–987.

[5] Bacchini, F., Ciammola, A., Iannaccone, R., and Marini, M. Combining forecasts for a flash estimate of Euro area GDP. *Contributi Istat* No. 3, National Statistical Institute of Italy, Rome, 2010.

[6] Bacon, D. W., and Watts, D. G. Estimating the transition between two intersecting straight lines. *Biometrika 58*, 3 (1971), 525–534.

[7] Bates, J. M., and Granger, C. W. The combination of forecasts. *Journal of the Operational Research Society 20*, 4 (1969), 451–468.

[8] Bekiroglu, K., Duru, O., Gulay, E., Su, R., and Lagoa, C. Predictive analytics of crude oil prices by utilizing the intelligent model search engine. *Applied Energy 228* (2018), 2387–2397.

[9] Bordignon, S., Bunn, D. W., Lisi, F., and Nan, F. Combining day-ahead forecasts for british electricity prices. *Energy Economics 35* (2013), 88–103.

[10] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, 2015.

[11] Caldeira, J. F., Moura, G. V., Nogales, F. J., and Santos, A. A. Combining multivariate volatility forecasts: an economic-based approach. *Journal of Financial Econometrics 15*, 2 (2017), 247–285.

[12] Chen, Z.-Y. Integration learning of neural network training with swarm intelligence and meta-heuristic algorithms for spot gold price forecast. *Applied Artificial Intelligence 36*, 1 (2022), 1994217.

[13] Clark, T. E., and McCracken, M. W. Combining forecasts from nested models. *Oxford Bulletin of Economics and Statistics 71*, 3 (2009), 303–329.

[14] Clemen, R. T. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting 5*, 4 (1989), 559–583.

[15] Cybenko, G. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems 2*, 4 (1989), 303–314.

[16] De Menezes, L. M., Bunn, D. W., and Taylor, J. W. Review of guidelines for the use of combined forecasts. *European Journal of Operational Research 120*, 1 (2000), 190–204.

[17] Dean, N. E., y Piontti, A. P., Madewell, Z. J., Cummings, D. A. T., Hitchings, M. D. T., Joshi, K., Kahn, R., Vespignani, A., Halloran, M. E., and Longini Jr, I. M. Ensemble forecast modeling for the design of Covid-19 vaccine efficacy trials. *Vaccine 38*, 46 (2020), 7213–7216.

[18] Dewangan, C. L., Singh, S. N., and Chakrabarti, S. Combining forecasts of day-ahead solar power. *Energy 202* (2020), 117743.

[19] Di Gangi, L. Sparse convex combinations of forecasting models by meta learning. *Expert Systems with Applications 200* (2022), 116938.

[20] Gardner Jr, E. S. Exponential smoothing: The state of the art–Part II. *International Journal of Forecasting 22*, 4 (2006), 637–666.

[21] Genre, V., Kenny, G., Meyler, A., and Timmermann, A. Combining expert forecasts: Can anything beat the simple average? *International Journal of Forecasting 29*, 1 (2013), 108–121.

[22] Graefe, A., Armstrong, J. S., Jones Jr, R. J., and Cuzán, A. G. Combining forecasts: An application to elections. *International Journal of Forecasting 30*, 1 (2014), 43–54.

[23] Gulay, E., and Duru, O. Hybrid modeling in the predictive analytics of energy systems and prices. *Applied Energy 268* (2020), 114985.

[24] Hibon, M., and Evgeniou, T. To combine or not to combine: selecting among forecasts and their combinations. *International Journal of Forecasting 21*, 1 (2005), 15–24.

[25] Hipel, K. W., and McLeod, A. I. *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier, 1994.

[26] Hong, H., Kubik, J. D., and Solomon, A. Security analysts' career concerns and herding of earnings forecasts. *The RAND Journal of Economics 31*, 1 (2000), 121–144.

[27] Huang, H., and Lee, T.-H. To combine forecasts or to combine information? *Econometric Reviews 29*, 5-6 (2010), 534–570.

[28] Hyndman, R., Koehler, A. B., Ord, J. K., and Snyder, R. D. *Forecasting with Exponential Smoothing: The State Space Approach*. Springer, Berlin, 2008.

[29] Hyndman, R. J., Koehler, A. B., Snyder, R. D., and Grose, S. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting 18*, 3 (2002), 439–454.

[30] Jiang, P., Liu, Z., Niu, X., and Zhang, L. A combined forecasting system based on statistical method, artificial neural networks, and deep learning methods for short-term wind speed forecasting. *Energy 217* (2021), 119361.

[31] Jose, V. R. R., and Winkler, R. L. Simple robust averages of forecasts: Some empirical results. *International Journal of Forecasting 24*, 1 (2008), 163–169.

[32] Kourentzes, N. Intermittent demand forecasts with neural networks. *International Journal of Production Economics 143*, 1 (2013), 198–206.

[33] Lemke, C., and Gabrys, B. Meta-learning for time series forecasting and forecast combination. *Neurocomputing 73*, 10-12 (2010), 2006–2016.

[34] Lin, Y., Koprinska, I., Rana, M., and Troncoso, A. Solar power forecasting based on pattern sequence similarity and meta-learning. In *Artificial Neural Networks and Machine Learning – ICANN 2020. 29th International Conference on Artificial Neural Networks, Bratislava, Slovakia, September 15–18, 2020, Proceedings, Part I* (Cham, 2020), I. Farkaš, P. Masulli and S. Wermter, Eds., vol. 12396 of *Lecture Notes in Computer Science*, Springer, pp. 271–283.

[35] Luukkonen, R., Saikkonen, P., and Teräsvirta, T. Testing linearity against smooth transition autoregressive models. *Biometrika 75*, 3 (1988), 491–499.

[36] Makridakis, S., and Hibon, M. The M3-competition: results, conclusions and implications. *International Journal of Forecasting 16*, 4 (2000), 451–476.

[37] Marcjasz, G., Uniejewski, B., and Weron, R. Probabilistic electricity price forecasting with NARX networks: Combine point or probabilistic forecasts? *International Journal of Forecasting 36*, 2 (2020), 466–479.

[38] Ord, J. K., Koehler, A. B., and Snyder, R. D. Estimation and prediction for a class of dynamic nonlinear statistical models. *Journal of the American Statistical Association 92*, 440 (1997), 1621–1629.

[39] Palm, F. C., and Zellner, A. To combine or not to combine? Issues of combining forecasts. *Journal of Forecasting 11*, 8 (1992), 687–701.

[40] Priestley, M. B. Non-linear and Non-stationary Time Series Analysis. Academic Press, London, 1988.

[41] Rapach, D. E., Strauss, J. K., and Zhou, G. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies 23*, 2 (2010), 821–862.

[42] Reid, D. J. Combining three estimates of gross domestic product. *Economica 35*, 140 (1968), 431–444.

[43] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. Learning representations by back-propagating errors. *Nature 323*, 6088 (1986), 533–536.

[44] St-Aubin, P., and Agard, B. Precision and reliability of forecasts performance metrics. *Forecasting 4*, 4 (2022), 882–903.

[45] Stock, J. H., and Watson, M. W. Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting 23*, 6 (2004), 405–430.

[46] Subba Rao, T., and Gabr, M. *An Introduction to Bispectral Analysis and Bilinear Time Series Models*. Springer Science & Business Media, 2012

[47] Teräsvirta, T. Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association 89*, 425 (1994), 208–218.

[48] Teräsvirta, T., and Anderson, H. M. Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics 7*, 51 (1992), 119–136.

[49] Tong, H. *Non-linear time series: a dynamical system approach*. Oxford University Press, 1990.

[50] Tong, H., and Lim, K. S. Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society: Series B (Methodological) 42*, 3 (1980), 245–268.

[51] Velleman, P. F., and Hoaglin, D. C. *Applications, Basics, and Computing of Exploratory Data Analysis*. Duxbury Press, 1981.

[52] Wallis, K. F. Combining forecasts–forty years later. In *Perspectives on Econometrics and Applied Economics*. M. Taylor, Ed., Routledge, 2014, pp. 39–48.

[53] Woodward, W. A., Gray, H. L., and Elliott, A. Stationary time series. In *Applied Time Series Analysis*, W. A. Woodward, H. L. Gray and A. Elliott, Eds., CRC Press, 2011, pp. 25–76.

[54] Yamana, T. K., Kandula, S., and Shaman, J. Individual versus superensemble forecasts of seasonal influenza outbreaks in the United States. *PLoS computational biology 13*, 11 (2017), e1005801.

[55] Yang, H., Zhu, Z., Li, C., and Li, R. A novel combined forecasting system for air pollutants concentration based on fuzzy theory and optimization of aggregation weight. *Applied Soft Computing 87* (2020), 105972.

[56] Zhang, G., Patuwo, B. E., and Hu, M. Y. Forecasting with artificial neural networks:: The state of the art. *International Journal of Forecasting 14*, 1 (1998), 35–62.

[57] Zou, H. F., Xia, G. P., Yang, F. T., and Wang, H. Y. An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting. *Neurocomputing 70*, 16-18 (2007), 2913–2923.