

## Article

# Analysis of Weighting Strategies for Improving the Accuracy of Combined Forecasts

José V. Segura-Heras <sup>1,\*</sup>, José D. Bermúdez <sup>2</sup>, Ana Corberán-Vallet <sup>2</sup> and Enriqueta Vercher <sup>2</sup>

<sup>1</sup> I.U. Operations Research Center, University Miguel Hernandez of Elche, Avda. Ferrocarril s/n, 03202 Elche, Spain

<sup>2</sup> Department of Statistics and O.R., University of Valencia, C/ Dr. Moliner 50, 46100 Burjassot, Spain; jose.d.bermudez@uv.es (J.D.B.); ana.corberan@uv.es (A.C.-V.); enriqueta.vercher@uv.es (E.V.)

\* Correspondence: jvsh@umh.es

**Abstract:** This paper deals with the weighted combination of forecasting methods using intelligent strategies for achieving accurate forecasts. In an effort to improve forecasting accuracy, we develop an algorithm that optimizes both the methods used in the combination and the weights assigned to the individual forecasts, COMbEB. The performance of our procedure can be enhanced by analyzing separately seasonal and non-seasonal time series. We study the relationships between prediction errors in the validation set and those of ex-post forecasts for different planning horizons. This study reveals the importance of setting the size of the validation set in a proper way. The performance of the proposed strategy is compared with that of the best prediction strategy in the analysis of each of the 100,000 series included in the M4 Competition.

**Keywords:** forecasting; time series methods; forecasting combination; M4 Competition

**MSC:** 62-08; 62L05; 62P20



**Citation:** Segura-Heras, J.V.; Bermúdez, J.D.; Corberán-Vallet, A.; Vercher, E. Analysis of Weighting Strategies for Improving the Accuracy of Combined Forecasts. *Mathematics* **2022**, *10*, 725. <https://doi.org/10.3390/math10050725>

Academic Editor: Asier Ibeas

Received: 2 January 2022

Accepted: 23 February 2022

Published: 24 February 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In uncertain environments, decision making based on the analysis of historical data is of utter importance. This often implies the implementation and development of different prediction strategies, which have to be adapted to the specific characteristics of the data. Nowadays, the vast majority of companies have access to large datasets. Hence, the development of simple, accessible, competitive, and automatic tools that provide accurate forecasts in a reasonable computing time is a need.

It is well known that the combination of forecasts derived from methods that may differ substantially and draw from different sources of information can improve the forecasting accuracy in comparison with forecasts provided individually by these methods ([1–3], for example). For instance, [4] proposes using exponential weighted information criteria, while other authors use neural networks (see, for instance, [5–7]) and machine learning techniques ([8,9]) for improving the accuracy of the forecasts. The key is to note the importance of deciding on suitable and robust procedures to select the contributed models and their assigned weights so as to produce accurate out-of-sample point forecasts. To this end, this paper proposes a procedure for suitably selecting both methods and weights.

The M4 Competition [10], available in the R package M4comp2018 [11], is an open competition that includes a large number of time series. In particular, the package contains 100,000 series, each of them with the following components: the series, the true future values (the test part), type, and domain, and the submitted forecasts of the top 25 participants. The competition was used to establish a ranking of forecasting procedures based on different strategies (hybrid technologies, combination of prediction methods, statistical approaches, etc.) to build accurate forecasts. One of the main objectives of this competition was to improve the performance of forecasting methods that use historical data for decision

making (see, for instance, [12] and references therein). As shown, the most accurate results were obtained by implementing several exponential smoothing formulations within neural networks [13] and applying a meta-model to combine forecasts [14].

Within the framework of combining forecasts, our objective in this paper is to develop a procedure to select both forecasting methods and their corresponding weights based on the behavior of well-established models and weighting strategies. To do so, we work with the time series of the M4 Competition under the same conditions set with respect to the seasonal pattern (yearly, quarterly, monthly, and other time series) and forecasting horizons (6 years, 8 quarters, 18 months, 14 days, 13 weeks, and 48 h, depending on the nature of the series). In particular, we develop an experiment in which each time series is segmented into two subsets: the training set and the validation set, which contains the last observations and whose size coincides with that of the established forecasting horizon. The performance of several model selection and weighting strategies is compared using the yearly and quarterly time series. This involves the analysis of the forecasting results of 47,000 time series. Based on this comparison, we then present a new algorithm for combining forecasts, COMBEB. The performance of our procedure is assessed using the remaining 53,000 series from the competition.

This paper is organized as follows. In the next section, we briefly review the forecasting methods that we use in the convex linear combination and several weighting criteria. The strategies used in the experiment for the selection of models and weights are also developed in Section 2. Section 3 presents a new procedure to combine the selected forecasting methods. The main results obtained by applying the new algorithm in the forecasting of the time series classified as either monthly or other time series are shown in Section 4. The last section offers some concluding remarks.

## 2. Materials and Methods

It is well known that the linear combination of different forecasting methods may produce more accurate out-of-sample forecasts. Within this scenario, the computation of sensible weights also plays an important role, and different strategies have been proposed in an effort to improve the forecasting accuracy (see [1], and references therein). One of the most widely used error-based weighting strategies is the one proposed in [2]. Here, the weight assigned to each method is inversely proportional to its forecasting MSE. This strategy has also been used with other measures of the forecasting error. Alternatively, many authors consider convex combinations of forecasts, that is, linear combinations in which the weights are constrained to be non-negative and add up to one. This simplifies the interpretation of the weights.

In the next subsection, we briefly describe the forecasting methods and the weighting strategies that we use in our combination.

### 2.1. Forecasting Methods

Let  $Y = \{y_1, y_2, \dots, y_N\}$  be the time series under study and  $p$  be the length of the seasonal cycle. At time  $N$ , we want to forecast the vector of future observations  $\hat{Y} = \{\hat{y}_{N+1}, \hat{y}_{N+2}, \dots, \hat{y}_{N+h}\}$ ,  $h$  being the forecasting horizon. Forecasts for the time series are calculated by using the following models, which can be grouped into three categories:

Category 1: Those models that use the last observations to provide forecasts.

- Naïve 1 method (*Naïve*). The forecast coincides with the last observation:  $\hat{y}_t = y_{t-1}$ .
- Naïve 2 method (*Naïve2*). The forecast includes an estimation of the trend:  $\hat{y}_t = y_{t-1} + (y_{t-1} - y_{t-2})$ .
- Seasonal Naïve method (*SNaïve2*). The forecast is the observation in the same period of the previous seasonal cycle:  $\hat{y}_t = y_{t-p}$ .
- Moving average of  $k$  periods (*MAk*). The forecast is given by the arithmetic average of the last  $k \geq 2$  observations:  $\hat{y}_t = \frac{y_{t-1} + y_{t-2} + \dots + y_{t-k}}{k}$ .

Category 2: Exponential smoothing models (see, for instance, Refs. [15,16]) which can analyze either the original data (named raw data) or their logarithmic transformation (named ln data). Note that the logarithmic transformation converts the multiplicative trend and seasonal effects into additive ones. To work with exponential smoothing models, the initial values for the level, trend, seasonal effects, and the smoothing parameters must be estimated using the observed data. In our approach, the corresponding initial values are also considered as parameters of the model, and they are jointly estimated using statistical and optimization tools (see, for instance, Refs. [17–19]). In particular:

- For the analysis of non-seasonal time series, we apply Gardner’s damped trend model ([20,21]), which is denoted as *G-Raw data* or *G-ln data*, indicating the model (Gardner) and the data analyzed (raw or log transformed), respectively.
- For the analysis of seasonal time series, the additive Holt–Winters model with a damped trend is applied ([18,22]). Analogously, the results associated with this model are respectively identified as *HW-Raw data* or *HW-ln data*.

Category 3: Finally, in order to incorporate the advantages of ARIMA models [23], we consider the forecasts provided by using the *auto.arima* function included in the forecast package of R [24], which is available in CRAN (<http://cran.r-project.org/> accessed on 15 February 2022). See, also, Ref. [25].

### 2.2. Linear Combination of Forecasting Techniques

In this paper, we consider a two-stage procedure for combining forecasting models. In a first stage,  $h$  forecasts are collected for every time series  $Y$  individually, using the aforementioned models,  $M_j$ , for  $j = 1, \dots, 9$ . Since these take into account the value of the parameter of the seasonal cycle,  $p$ , only six models can be considered either for seasonal or non-seasonal time series. In the second stage, the procedure computes a linear convex combination of the selected models with their assigned weights  $w_j$  to provide the new  $h$  ex-post combined forecasts for each time series:  $\hat{Y}$ .

Let us now analyze three error-based weighting strategies. They are based on either the fitting sMAPE or the forecasting sMAPE, depending on the data used to compute the weights.

First, the procedure makes a segmentation of the observed data set  $Y = \{y_1, \dots, y_N\}$  into two subsets: the training set, which contains the first observations,  $T_1 = \{y_1, \dots, y_{N-h}\}$ , and the validation subset  $T_2$ , with the last  $h$  observations. Using  $T_1$ , each  $M_j$  method provides  $h$  forecasts:  $\hat{T}_2^j = \{\hat{y}_{N-h+1}^j, \dots, \hat{y}_N^j\}$ . The fitting and forecasting sMAPE errors for each method are respectively evaluated as follows:

$$sMAPE_{T_1}^j = \frac{200}{N-h} \sum_{t=1}^{N-h} \frac{|y_t - \hat{y}_t^j|}{|y_t + \hat{y}_t^j|}, \tag{1}$$

and

$$sMAPE_{T_2}^j = \frac{200}{h} \sum_{t=N-h+1}^N \frac{|y_t - \hat{y}_t^j|}{|y_t + \hat{y}_t^j|}. \tag{2}$$

The  $k$  ( $2 \leq k \leq 6$ ) models selected to be included in the linear convex combination are those with smaller forecasting sMAPE ( $sMAPE_{T_2}^j$ ). The value of  $k$  is set by the forecaster. The weights assigned to the methods included in the combination can be determined by means of one of the following strategies:

- *sMAPE* error-based (EB). The weights  $w_j$  are the normalized inverse of the sMAPE fitting errors obtained in  $T_1$  for each model, that is:

$$w_j = \frac{1}{sMAPE_{T_1}^j} \bigg/ \sum_{i=1}^k \left( \frac{1}{sMAPE_{T_1}^i} \right), \quad j = 1, \dots, k \tag{3}$$

for the  $k$  methods included in the linear combination.

- Optimal weights for time series (OW-ts). For the time series  $Y$ , the optimal weights assigned to each method,  $w_j$ , correspond to the solution of the following nonlinear optimization problem:

$$\begin{aligned} \text{Min } sMAPE_{T_2}^C &= \frac{200}{h} \sum_{t=N-h+1}^N \frac{|y_t - \sum_{j=1}^k w_j \hat{y}_t^j|}{|y_t + \sum_{j=1}^k w_j \hat{y}_t^j|} \\ \text{s.t. } \sum_{j=1}^k w_j &= 1, \quad w_j \geq 0, \quad j = 1, \dots, k \end{aligned} \tag{4}$$

- Optimal weights for time series and period (OW-tsp). The weights are dynamically obtained depending on both the time series and the position within the forecasting horizon,  $i = 1, \dots, h$ . For the time series  $Y$ , the optimal weights assigned to each method are given by the  $k \times h$  vector solution of the following nonlinear optimization problem:

$$\begin{aligned} \text{Min } sMAPE_{T_2}^{Ch} &= \frac{200}{h} \sum_{t=N-h+1}^N \frac{|y_t - \sum_{j=1}^k w_{j,i} \hat{y}_t^j|}{|y_t + \sum_{j=1}^k w_{j,i} \hat{y}_t^j|} \\ \text{s.t. } \sum_{j=1}^k \sum_{i=1}^h w_{j,i} &= 1, \\ w_{j,i} &\geq 0, \quad j = 1, \dots, k \quad i = 1, \dots, h \end{aligned} \tag{5}$$

$i = t + h - N$  being the position of the forecast within the forecasting horizon; that is,  $i = 1$  when  $t = N - h + 1$ ;  $i = 2$  when  $t = N - h + 2$ , and so on.

The optimization problems defined in (4) and (5) are solved using simulation techniques. The procedure randomly generates from 5000 to 10,000 vectors of weights, using the Dirichlet distribution with all its parameters equal to 1, which is equivalent to the uniform distribution in the  $(k - 1)$ -simplex, for the different values of  $k = 2, \dots, 6$ . For each time series, it chooses the value of  $k$  and the weights that obtain the minimum of the objective function  $sMAPE_{T_2}^C$  or  $sMAPE_{T_2}^{Ch}$ , depending on the strategy used.

Finally, in the second stage, using the complete time series  $Y$ ,  $h$ -steps ahead forecasts (ex-post forecasts) are calculated individually using the  $k$  forecasting models selected:  $\hat{Y}^j = \{\hat{y}_t^j\}_{t=N+1}^{N+h}$ . Regarding the calculation of the weights at this stage, there is a small difference. For the weighting strategies OW-ts and OW-tsp, those weights previously determined as a solution of the corresponding optimization problems are applied. Instead, for the sMAPE error-based strategy, the weights are recalculated using all the data in  $Y$ , as follows:

$$w_j = \frac{\frac{1}{sMAPE_Y^j}}{\sum_{i=1}^k \left(\frac{1}{sMAPE_Y^i}\right)}, \quad j = 1, \dots, k \tag{6}$$

Then, the weights are applied to build the convex linear combination of the  $h$  ex-post forecasts:

$$\hat{Y}^C = \{\hat{y}_{N+1}^c, \dots, \hat{y}_{N+h}^c\}, \tag{7}$$

where  $\hat{y}_t^c = \sum_{j=1}^k w_j \hat{y}_t^j$ , for  $t = N + 1, \dots, N + h$ , and  $w_j$  ( $w_{j,i}$ , within OW-tsp) is the non-negative weight assigned to the  $j$ th forecasting method.

Our objective is to select an appropriate set of weights for a subset of models, so that the resultant combined forecasts are the most accurate ex-post forecasts. Let us present an experiment that is carried out to both assess the three weighting strategies previously described and provide guidance for the selection of  $k$  (the number of methods included in the linear combination).

### 2.3. Model and Weighting Selection: An Experiment

The experiment works as follows: First, forecasts for  $T_2$  are individually obtained using six forecasting models and the data in  $T_1$ . For the non-seasonal time series, we consider: Naive, Naive2, MA3, G-Raw data, G-In data and ARIMA models. For the seasonal series: Naive, SNaive2, MA3, HW-Raw data, HW-In data, and ARIMA. The fitting and forecasting sMAPE for  $T_1$  and  $T_2$  are obtained for each model. Then, it defines  $m$  different linear combinations of these forecasting methods,  $LC_m$  for  $m = 1, \dots, 16$ . Gardner’s damped trend and Holt–Winters model are always included in the corresponding linear combination. The corresponding weights, for every combination, have been calculated using the three proposed weighting strategies: EB, OW-ts, and OW-tsp.

The experiment has been tested on two sets of time series from the M4 Competition, the yearly and quarterly data sets ([10]). The forecasting  $sMAPE_{T_2}^C$  are used to analyze both the effect of the cardinality of the linear convex combination and the performance of the weighting strategies.

#### 2.3.1. Yearly Time Series

Table 1 shows the fitting and forecasting sMAPE obtained by individually applying Naive, Naive2, MA3, G-Raw data, G-In data, and ARIMA models for the the set of non-seasonal yearly time series, whose prediction horizon is  $h = 6$ . The lowest average forecasting error in  $T_2$  was obtained using the automatic ARIMA with 15.22% of sMAPE.

**Table 1.** Average sMAPE for the 23,000 yearly time series, for  $h = 6$ .

sMAPE	Naive	Naive2	MA3	G-Raw Data	G-In Data	ARIMA
$T_1$	8.84	9.41	13.47	7.47	6.73	6.44
$T_2$	19.25	22.60	21.48	16.30	16.72	15.22

The combined forecasting errors obtained in  $T_2$  by using each specified linear combination are shown in Table 2. The horizontal lines between rows enable us to compare the results with respect to the cardinality of the linear combination. As expected, combined forecasting usually provides more accurate forecasts than those obtained with an individual forecasting model. Practically all the combinations obtain lower forecasting errors in  $T_2$ . It is also worth noting that increasing the cardinality in the convex combination does not clearly improve the accuracy, as indicated in [1].

**Table 2.** Average  $sMAPE^C$  in the validation set  $T_2$  for the yearly time series, using 16 linear combinations and the three weighting strategies: EB, OW-ts, and OW-tsp. In bold the minimum  $sMAPE^C$ .

Linear Combination ( $LC_m, m = 1, \dots, 16$ )	EB	OW-ts	OW-tsp
G-Raw data, G-In data	16.13	14.18	13.55
G-Raw data, G-In data, Naive	15.59	11.84	10.39
G-Raw data, G-In data, Naive2	14.70	9.48	7.09
G-Raw data, G-In data, MA3	15.98	12.49	11.14
G-Raw data, G-In data, ARIMA	14.98	11.62	10.29
G-Raw data, G-In data, Naive, Naive2	14.68	9.02	6.69
G-Raw data, G-In data, Naive, MA3	15.81	11.84	10.35
G-Raw data, G-In data, Naive, ARIMA	14.88	10.79	9.22
G-Raw data, G-In data, Naive2, MA3	14.70	8.97	6.36
G-Raw data, G-In data, Naive2, ARIMA	14.44	9.14	6.82
G-Raw data, G-In data, MA3, ARIMA	15.04	11.01	9.37
G-Raw data, G-In data, Naive, Naive2, MA3	14.85	8.90	6.46
G-Raw data, G-In data, Naive, Naive2, ARIMA	14.48	8.80	6.41
G-Raw data, G-In data, Naive, MA3, ARIMA	15.07	10.75	9.12
G-Raw data, G-In data, Naive2, MA3, ARIMA	14.49	8.83	6.32
G-Raw data, G-In data, Naive, Naive2, MA3, ARIMA	14.61	8.95	6.61

On the other hand, there seems to be large differences between the average forecasting sMAPE when the strategies based on optimal weights are used. Are these strategies overfitting the data? To analyze this question, we have also calculated the *ideal* average sMAPE in the validation subset  $T_2$ , which is the forecasting  $sMAPE_{T_2}$  that we would have obtained if we had chosen the method that provided the minimum forecasting error for every time series. For the set of yearly series, this *ideal* sMAPE error was 9.52% in  $T_2$ , confirming that the overfitting problem has arisen.

### 2.3.2. Quarterly Time Series

We now describe the performance of our experiment for the 24,000 quarterly time series from the M4 Competition, the seasonal cycle being  $p = 4$ , and the forecasting horizon  $h = 8$ . Table 3 shows the averaged fitting and forecasting sMAPE obtained by individually applying Naive, SNaive2, MA3, HW-Raw data, HW-In data, and ARIMA methods. The lowest averaged forecasting error in  $T_2$  was obtained using the HW-In data with 10.97% of sMAPE.

**Table 3.** Average sMAPE for the 24,000 quarterly time series, for  $h = 8$ . In bold the minimum sMAPE at  $T_2$ .

sMAPE	Naive	SNaive2	MA3	HW-Raw Data	HW-In Data	ARIMA
$T_1$	8.20	11.96	10.38	6.92	6.75	6.70
$T_2$	11.85	12.82	11.82	11.03	10.97	11.05

Table 4 shows a summary of the averaged forecasting  $sMAPE^C$  in  $T_2$  for the 16 specified convex linear combinations and weighting strategies. Overall, we see that increasing the cardinality of the linear combination does not necessarily imply a better accuracy. So, if we decide to apply the EB weighting strategy (in order to avoid the problem of overfitting), the lowest  $sMAPE^C$  (10.37%) is obtained using the combination of four models *HW-Raw data, HW-In data, Naive, and ARIMA*.

**Table 4.** Average  $sMAPE^C$  in the validation set  $T_2$ , for the quarterly time series, using 16 linear combinations and three weighting strategies: EB, OW-ts, and OW-tsp. In bold the minimum  $sMAPE^C$ .

Linear Combination ( $LC_m$ )	EB	OW-ts	OW-tsp
HW-Raw data, HW-In data	10.84	10.23	9.83
HW-Raw data, HW-In data, Naive	10.54	8.81	7.53
HW-Raw data, HW-In data, SNaive2	10.87	9.36	7.88
HW-Raw data, HW-In data, MA3	10.70	9.25	8.14
HW-Raw data, HW-In data, ARIMA	10.47	8.85	7.72
HW-Raw data, HW-In data, Naive, SNaive2	10.62	8.72	7.36
HW-Raw data, HW-In data, Naive, MA3	10.57	8.72	7.36
HW-Raw data, HW-In data, Naive, ARIMA	10.37	8.31	6.87
HW-Raw data, HW-In data, SNaive2, MA3	10.77	9.06	7.36
HW-Raw data, HW-In data, SNaive2, ARIMA	10.49	8.43	6.69
HW-Raw data, HW-In data, MA3, ARIMA	10.43	8.46	7.10
HW-Raw data, HW-In data, Naive, SNaive2, MA3	10.65	8.58	7.51
HW-Raw data, HW-In data, Naive, SNaive2, ARIMA	10.41	8.10	7.03
HW-Raw data, HW-In data, Naive, MA3, ARIMA	10.40	8.24	7.35
HW-Raw data, HW-In data, SNaive2, MA3, ARIMA	10.48	8.28	7.22
HW-Raw data, HW-In data, Naive, SNaive2, MA3, ARIMA	10.44	8.16	6.40

### 2.3.3. Comparisons with Ex-Post Forecasts

The M4 Competition provides the values of the true future values  $T_{new} = \{y_{N+1}, \dots, y_{N+h}\}$  [11]. Therefore, it is possible to calculate the ex-post forecasting errors,  $e_t = y_t - \hat{y}_t^c$ , for  $t = N + 1, \dots, N + h$ , for all linear combinations and for the three weighting strategies, allowing us to analyze their ex-post performance. Here, we show these results for the linear combinations of four methods.

Tables 5 and 6 show the averaged forecasting  $sMAPE^C$  for the yearly and quarterly time series, respectively. For yearly time series, the best result is obtained for the combination *G-Raw data, G-In data, Naive, and ARIMA*, with the weights provided by the EB strategy, with a  $sMAPE^C$  of 13.66%. These four methods also obtained the lowest forecasting  $sMAPE$  in  $T_2$ , when they were individually applied to the yearly time series. For quarterly series, the best result is obtained with the combination *HW-Raw data, HW-In data, Naive, and ARIMA*. It is worth emphasizing here that if we select the  $k = 4$  methods with the smallest forecasting  $sMAPE$  in  $T_2$ , we achieve a very accurate combined ex-post forecast.

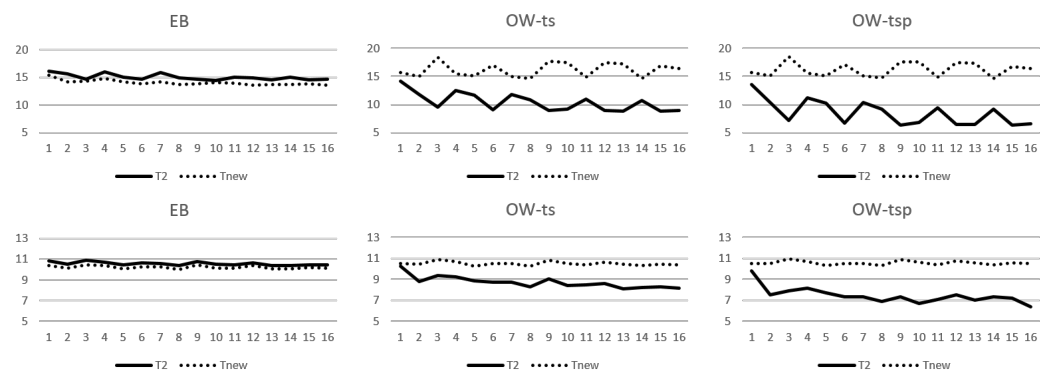
In addition, the forecasting errors obtained in  $T_2$ , when the EB weighting strategy was applied, are close to the ones in  $T_{new}$ . However, this is not the case when the optimal and dynamic weighting strategies are applied. Figure 1 shows the comparison of the averaged forecasting  $sMAPE^C$  obtained in  $T_2$  and  $T_{new}$  for all the corresponding 16 linear convex combinations proposed for the yearly and quarterly time series and using the three weighting strategies. The overfitting clearly appears when OW-ts and OW-tsp weighting strategies are applied. Based on these results, we propose to use the  $sMAPE$  error-based (EB) weighting strategy, which is the simplest one and provides the best ex-post combined forecasts.

**Table 5.** Yearly time series: average  $sMAPE^C$  of the ex-post forecasts in  $T_{new}$ , using convex combinations of four forecasting methods and three weighting strategies. In bold the minimum  $sMAPE^C$ .

Linear Combination	EB	OW-ts	OW-tsp
G-Raw data, G-In data, Naive, Naive2	13.75	17.08	17.18
G-Raw data, G-In data, Naive, MA3	14.12	14.96	15.09
G-Raw data, G-In data, Naive, ARIMA	13.66	14.69	14.74
G-Raw data, G-In data, Naive2, MA3	13.83	17.77	17.58
G-Raw data, G-In data, Naive2, ARIMA	14.06	17.49	17.60
G-Raw data, G-In data, MA3, ARIMA	13.92	14.89	14.90

**Table 6.** Quarterly time series: average  $sMAPE^C$  of the ex-post forecasts in  $T_{new}$ , using linear combinations of four methods and three weighting strategies. In bold the minimum  $sMAPE^C$ .

Linear Combination	EB	OW-ts	OW-tsp
HW-Raw data, HW-In data, Naive, SNaive2	10.28	10.50	10.51
HW-Raw data, HW-In data, Naive, MA3	10.27	10.50	10.51
HW-Raw data, HW-In data, Naive, ARIMA	10.00	10.28	10.32
HW-Raw data, HW-In data, SNaive2, MA3	10.47	10.80	10.90
HW-Raw data, HW-In data, SNaive2, ARIMA	10.13	10.48	10.63
HW-Raw data, HW-In data, MA3, ARIMA	10.10	10.35	10.39



**Figure 1.** Forecasting  $sMAPE^C$  obtained in the validation set ( $T_2$ ) and  $T_{new}$  for all the proposed convex combinations  $LC_m$  ( $m = 1, \dots, 16$ ) and weighting strategies. The three upper graphs correspond to the yearly time series, while the three graphs at the bottom correspond to quarterly ones.

All calculations were performed on a workstation with two Intel® Xeon® E5-2650 v3 2.3 GHz processors with 10 physical cores and 20 threads (two processes per core). To carry out our study, we used the R language [26]. However, the estimation of parameters and predictions using Gardner's damped trend and Holt–Winters models were obtained with the SIOPRED tool ([19,21]).

### 3. Combination of Forecasts: A New Algorithm

The performance of the above experiment suggests using a reduced number of methods for combining forecasts and to apply the sMAPE error-based (EB) strategy to calculate the weights for the linear combination of these forecasts. Additionally, we propose selecting those models having the smallest sMAPE in the validation set  $T_2$  when applied individually. All these facts are included in the algorithm that we propose to combine forecasts, which we call COMbEB algorithm.

Let us assume we want to compute pointwise  $h$ -step-ahead forecasts for a set of  $S$  time series, all of them with the same seasonal cycle  $p$ . Our algorithm simultaneously analyzes the  $S$  series. In particular, it selects the  $k$  forecasting models, out of the  $J$  considered ones, with the smallest forecasting sMAPE in the validation set (when they are individually applied). Then, the sMAPE error-based (EB) weighting strategy is used to assign the weights to the models selected. Once the models and the weights have been set, the algorithm provides the ex-post combined forecasts. The outline of the algorithm is shown in Algorithm 1.

---

#### Algorithm 1 COMbEB

---

```

1: procedure COMBEB ALGORITHM( $S, Y_s, N_s, h, J, k$ )
2:   Let  $Y_s$  be a time series in  $S$  with length  $N_s$ . Divide  $Y_s$  into two subsets:  $T_{s,1} = \{y_1, \dots, y_{N_s-h}\}$  and  $T_{s,2} = \{y_{N_s-h+1}, \dots, y_{N_s}\}$ .
3:   Let  $M_j$  be one of the  $J$  considered forecasting methods.
4:   for all  $j \in J$  do
5:     for all  $s \in S$  do
6:       Apply the  $M_j$  method using  $T_{s,1}$  and calculate the  $h$ -step-ahead forecasts:  $\hat{y}_{s,t}^j$ ,  $t = N_s - h + 1, \dots, N_s$ .
7:       Apply Equation (2) to calculate  $sMAPE_{T_{s,2}}^j$ .
8:     end for
9:     Evaluate the average  $sMAPE_S^j$  for the  $S$  time series.
10:  end for
11:  Select the  $k \leq J$  models with the lowest average  $sMAPE_S^j$ , arranged in increasing order of forecasting error.
12:  if  $sMAPE_S^k \geq 2 sMAPE_S^1$  then
13:    Remove the  $k$ th model from this set of methods.
14:  end if
15:  Let  $K$  be the final subset of selected methods, ( $K \leq k$ ).
16:  for all  $s \in S$  do
17:    for all  $j \in K$  do.
18:      Apply the  $M_j$  method to fit  $Y_s$ .
19:      Determine the weight  $w_{s,j}$  using Equation (6) for the  $sMAPE_{Y_s}^j$ .
20:      Apply the  $M_j$  method to generate the  $h$ -step-ahead forecasts:  $\hat{y}_{s,t}^j$  for  $t = N_s + 1, \dots, N_s + h$ .
21:    end for
22:    Compute the combined forecasts:  $\hat{y}_{s,t}^c = \sum_{j \in K} w_{s,j} \hat{y}_{s,t}^j$  for  $t = N_s + 1, \dots, N_s + h$ .
23:  end for
24: end procedure

```

---



Additionally, when a large amount of time series with the same characteristics have to be forecasted, and the size of the validation set is  $h$ , the performance of the training experiment allows us to provide an estimation of the future average forecasting error.

Note that Algorithm 1 can also be used to predict an individual time series  $Y$ . In that case, we propose to select the  $k$  methods, from a set of  $J$  forecasting methods, with the lowest forecasting error in the validation set  $T_2$ , which must have the size of the forecasting horizon  $h$ . The weights assigned to the selected models are the ones provided by the sMAPE error-based strategy (EB). A flowchart of this forecasting process can be seen in Figure 2.

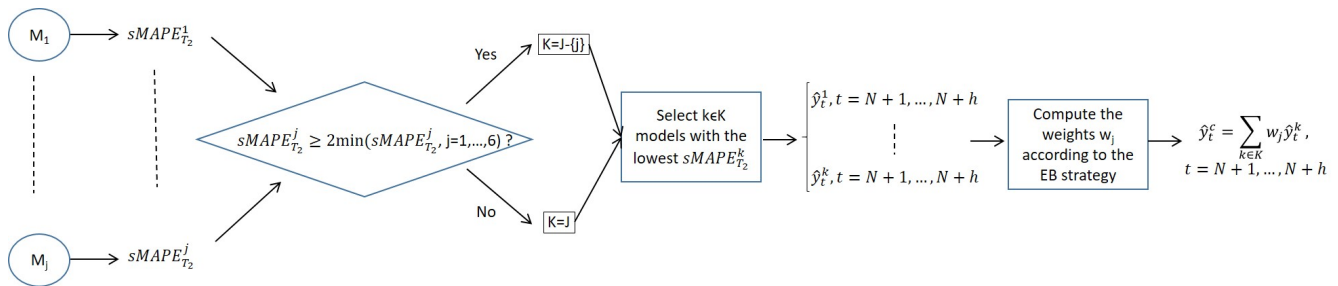


Figure 2. Flowchart of the COMbEB algorithm for an individual time series.

### 4. Numerical Results

In this section, we present the main results obtained with the COMbEB algorithm using the forecasting methods introduced in Section 2, and assuming  $J = 6$  and  $k = 4$ .

We use here the 100,000 time series of the M4 Competition [10]. Table 7 shows a summary of the main characteristics of its times series sets. It must be noted that the size of the validation set  $T_2$  will coincide with the forecasting horizon  $h$ , while the contributed forecasting methods depends on the presence of a seasonal cycle ( $p > 1$ ).

Table 7. Summary of the characteristics of the M4 Competition series.

Time Series	Yearly	Quarterly	Monthly	Weekly	Daily	Hourly
Size	23,000	24,000	48,000	359	4227	414
Seasonal pattern (p)	1	4	12	1	1	24
Forecasting horizon (h)	6	8	18	13	14	48

#### 4.1. Combined Forecasts for Non-Seasonal Time Series

For the yearly time series, the COMbEB algorithm selects the following four methods: *G-Raw data*, *G-ln data*, *Naive*, and *ARIMA* (see Table 1). Using these selected methods, the COMbEB algorithm determines the error-based weights and the individual ex-post  $h$ -step-ahead forecasts. Finally, it computes the combined forecasts for each time series. The average ex-post  $sMAPE^C$  is 13.66% (see Table 5).

For the weekly time series, the COMbEB algorithm selects the following methods: *G-ln data* (8.07%), *MA3* (8.17%), *G-Raw data* (8.46%), and *Naive* (9.2% of sMAPE in  $T_2$ ). The procedure generates the linear combination of their ex-post forecasts, whose average forecasting  $sMAPE^C$  is 9.02%.

For the daily time series, our algorithm selects the methods: *Naive* (2.79% of  $sMAPE_{T_2}$ ), *G-Raw data* (2.87%), *ARIMA* (2.87%), and *G-ln data* (2.88%), which obtained the lowest averaged forecasting error in the validation set  $T_2$ . The ex-post average  $sMAPE^C$  in  $T_{new}$  was 3.03%. Note that this is the same combination of methods that was selected for the non-seasonal set of yearly time series.

#### 4.2. Combined Forecasts for Seasonal Time Series

For the quarterly times series set, the linear combination is built using *HW-Raw data*, *HW-ln data*, *MA3*, and *ARIMA* models (the models with the smallest forecasting errors in  $T_2$ ;

see Table 3). For this combination, the COmbEB algorithm calculates an ex-post forecasting  $sMAPE^C$  of 10.10% (see Table 6).

For the monthly time series, Table 8 shows the average fitting and forecasting sMAPE. These results allow us to select the  $k = 4$  methods that will be used in the combination as explained in the COmbEB algorithm. In particular, the COmbEB algorithm selects the following four methods: *HW-In data*, *HW-Raw data*, *MA3*, and *ARIMA*, which have returned the best results in the validation set. Using these selected methods, the COmbEB algorithm determines the error-based weights and the individual ex-post  $h$ -step-ahead forecasts. Finally, it computes the combined forecasts for each monthly time series. The average ex-post  $sMAPE^C$  is 12.81%.

**Table 8.** Average sMAPE for the 48,000 monthly time series, for  $h = 18$ . In bold the minimum sMAPE.

sMAPE	Naive	SNaive2	MA3	HW-Raw Data	HW-In Data	ARIMA
$T_1$	7.79	15.53	8.35	6.17	6.08	6.61
$T_2$	13.99	15.18	13.05	12.58	12.44	13.60

For the hourly time series, the algorithm computes the  $sMAPE_{T_2}^j$  and selects the following models for the linear combination: *SNaive2* (14.57% of  $sMAPE^C$  in  $T_2$ ), *HW-Raw data* (15.96%), *HW-In data* (17.59%), and *ARIMA* (30.19%), respectively. At this point, since the averaged sMAPE obtained by the automatic *ARIMA* is more than twice the average sMAPE for the *SNaive2* method, the procedure allows us to decide whether or not to include the fourth forecasting method, especially considering the relationship between this forecasting error in  $T_2$  and the ex-post forecasting error.

In order to analyze the role of this restriction affecting the cardinality of the linear combination, we run the COmbEB algorithm twice: for a linear combination of four (*HW-Raw data*, *HW-In data*, *SNaive2*, *ARIMA*) and for a linear combination of three (*HW-Raw data*, *HW-In data*, *SNaive2*) forecasting models. Table 9 shows the average forecasting  $sMAPE^C$  obtained for these selected EB-weighted linear combinations, in both the validation and the prediction set,  $T_2$  and  $T_{new}$ . These results also show the importance of the information provided by the forecasting errors in the validation set, which provides an estimate of the future forecasting error, on average.

**Table 9.** Average forecasting sMAPE for the 4227 hourly time series, using two alternative linear combinations of methods. In bold the minimum  $sMAPE^C$ .

Linear Combination	$sMAPE_{T_2}^C$	$sMAPE_{T_{new}}^C$
HW-Raw data, HW-In data, SNaive2	15.62	13.46
HW-Raw data, HW-In data, SNaive2, ARIMA	17.37	14.96

### 4.3. Comparative Results

Finally, we compare our forecasting results with those published at the end of the competition. Global Winner refers to the results provided in [13], which apply a hybrid approach based on exponential smoothing methods and a neural network. This was the most accurate proposal for forecasting the 100,000 times series, on average. The FFORMA procedure automatically computes combined forecasts based on features of the time series. It achieved the second overall position [14]. The last row includes the benchmark used in the M4 Competition, an arithmetic mean combination of three models: simple, Holt, and Damped exponential smoothing [10].

Table 10 shows the average sMAPE for the different categories and the average for all of them. In brackets, we show the ratio (in %) between the best forecasting result in the M4 and the result with the other procedures. For the entire set of the M4 Competition series, the averaged sMAPE using the CombEB algorithm was 11.93%, which would

have hypothetically achieved 8th place in the M4 competition, performing better than the benchmark.

**Table 10.** Comparison of the average forecasting ex-post sMAPE.

Algorithms	Yearly	Quarterly	Monthly	Weekly	Daily	Hourly	Average
Global Winner	13.18 (100%)	9.68 (100%)	12.13 (100%)	7.82 (84%)	3.17 (77%)	9.33 (95%)	11.37
FFORMA	13.53 (97%)	9.73 (99%)	12.64 (96%)	7.63 (86%)	3.10 (79%)	11.51 (77%)	11.72
COmbEB	13.66 (96%)	10.10 (96%)	12.81 (95%)	9.02 (73%)	3.03 (81%)	13.46 (66%)	11.93
Benchmark	14.85 (89%)	10.18 (95%)	13.43 (90%)	8.94 (74%)	2.98 (82%)	22.05 (40%)	12.56

## 5. Conclusions

In an effort to improve forecast accuracy, we have tried to bring new ideas and strategies in the field of combining forecasts. Since the proposed algorithm can be easily adapted and implemented, decision-making platforms can benefit from these results.

In particular, we have presented a nonsophisticated competitive procedure, the CombEB algorithm, which provides accurate out-of-sample forecasts based on the convex combination of simple and accessible forecasting methods. The selection of methods is obtained using an error-based strategy working with the forecasting errors in the validation set, whose size should coincide with that of the forecasting horizon. An extensive computational analysis has been performed to assess the performance of the proposed procedure. In particular, we have carried out an experiment with the 100,000 time series from the M4 Competition.

The results show that applying dynamic and optimal strategies for calculating the weights does not benefit the accuracy of the ex-post forecasts. Therefore, we recommend using weights that are inversely proportional to the fitting sMAPE.

The number and nature of the methods that will be included in the combination may depend on the preferences of the forecasting researcher. However, a small number of well-known and accessible methods and a simple weighting strategy have produced good forecasting results. It is important to highlight the good behavior of the exponential smoothing models that were included in all the combinations.

In order for the algorithm to be implemented automatically by any user, it would be interesting to adapt it so that all the prediction models considered are implemented in R. The development of an R package that requires as input the time series to be studied (or the set of time series), the forecast horizon, and the number of models to be used in the combination would be of great interest from a practical viewpoint.

**Author Contributions:** Conceptualization, All authors; Formal analysis, J.D.B. and E.V.; Methodology, All authors; Software, J.V.S.-H. and J.D.B.; Writing—original draft, E.V.; Writing—review and editing, J.D.B., A.C.-V., J.V.S.-H. and E.V. All authors have read and agreed to the published version of the manuscript.

**Funding:** The authors thank the financial support from the Spanish Ministry for Economy and Competitiveness (Ministerio de Economía, Industria y Competitividad) under grant MTM2017-83850-P. José V. Segura-Heras acknowledges the financial support from the Generalitat Valenciana under project PROMETEO/2021/063.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Armstrong, J.S. *Principles of Forecasting*; Kluwer: Amsterdam, The Netherlands, 2001.
2. Bates, J.; Granger, C. The combination of forecasts. *Oper. Res. Q.* **1969**, *20*, 451–468. [[CrossRef](#)]
3. Clemen, R.T. Combining Forecasts: A review and annotated bibliography. *Int. J. Forecast.* **1989**, *5*, 559–583. [[CrossRef](#)]
4. Taylor, J.W. Exponential weighted information criteria for selecting among forecasting methods. *Int. J. Forecast.* **2008**, *24*, 513–524. [[CrossRef](#)]

5. Martins, V.L.M.; Werner, L. Forecast combination in industrial series: A comparison between individual forecasts and its combination with and without correlated errors. *Expert Syst. Appl.* **2012**, *39*, 11479–11486. [[CrossRef](#)]
6. Shi, S.; Da Xu, L.; Liu, B. Improving the accuracy of nonlinear combined forecasting using neural networks. *Expert Syst. Appl.* **1999**, *16*, 49–54.
7. dos Santos, R.D.O.V.; Vellasco, M.M. Neural expert weighting: A NEW framework for dynamic forecast combination. *Expert Syst. Appl.* **2015**, *42*, 8625–8636. [[CrossRef](#)]
8. Lemke, C.; Gabrys, B. Meta-learning for time series forecasting and forecasting combination. *Neurocomputing* **2010**, *73*, 2006–2016. [[CrossRef](#)]
9. Zou, H.; Yang, Y. Combining time series models for forecasting. *Int. J. Forecast.* **2004**, *29*, 69–84. [[CrossRef](#)]
10. Makridakis, S.; Spiliotis, E.; Assimakopoulos, V. The M4 Competition: Results, findings, conclusion and way forward. *Int. J. Forecast.* **2018**, *34*, 802–808. [[CrossRef](#)]
11. Montero-Manso, P.; Netto, C.; Talagala, C. *M4comp2018: Data from the M4-Competition*; R Package Version 0.2.0. 2018. Available online: [https://github.com/carlanetto/M4comp2018/releases/download/0.2.0/M4comp2018\\_0.2.0.tar.gz](https://github.com/carlanetto/M4comp2018/releases/download/0.2.0/M4comp2018_0.2.0.tar.gz) (accessed on 2 January 2022).
12. Makridakis, S.; Hibon, M. The M3-Competition: Results, conclusions and implications. *Int. J. Forecast.* **2000**, *16*, 451–476. [[CrossRef](#)]
13. Smyl, S. A hybrid method of exponential smoothing and recurrent neural networks for time series Forecasting. *Int. J. Forecast.* **2020**, *36*, 75–85. [[CrossRef](#)]
14. Montero-Manso, P.; Athanasopoulos, G.; Hyndman, R.J.; Talagata, T.S. FFORMA: Feature-based forecast model averaging. *Int. J. Forecast.* **2020**, *36*, 86–92. [[CrossRef](#)]
15. Gardner, E.S., Jr. Exponential smoothing: The state of the art-part II. *Int. J. Forecast.* **2006**, *22*, 637–666. [[CrossRef](#)]
16. Hyndman, R.J.; Koehler, A.B.; Ord, J.K.; Snyder, R.D. *Forecasting with Exponential Smoothing. The State Space Approach*; Springer: Berlin, Germany, 2008.
17. Bermúdez, J.D.; Segura, J.V.; Vercher, E. Improving demand forecasting accuracy using non-linear programming software. *J. Oper. Res. Soc.* **2006**, *57*, 94–100. [[CrossRef](#)]
18. Bermúdez, J.D.; Segura, J.V.; Vercher, E. Holt-Winters forecasting: An alternative formulation applied to UK air passenger data. *J. Appl. Stat.* **2007**, *34*, 1075–1090. [[CrossRef](#)]
19. Bermúdez, J.D.; Segura, J.V.; Vercher, E. SIOPRED: A prediction and optimisation integrated system for demand. *Top* **2008**, *16*, 258–271.
20. Gardner, E.S., Jr.; McKenzie, E. Forecasting trends in time series. *Manag. Sci.* **1985**, *31*, 1237–1246. [[CrossRef](#)]
21. Vercher, E.; Corberán-Vallet, A.; Segura, J.V.; Bermúdez, J.D. Initial conditions estimations for improving forecast accuracy in exponential smoothing. *Top* **2012**, *20*, 517–533. [[CrossRef](#)]
22. Chatfield, C.; Yar, M. Holt-Winters forecasting: Some practical issues. *Statistician* **1988**, *37*, 129–140. [[CrossRef](#)]
23. Box, G.; Jenkins, G. *Time Series Analysis: Forecasting and Control*; Holden-Day: San Francisco, CA, USA, 1976.
24. Hyndman, R.; Khandakar, Y. Automatic time series forecasting: The forecast package for R. *J. Stat. Softw.* **2008**, *26*, 1–22.
25. Hyndman, R.; Athanasopoulos, G.; Bergmeir, C.; Caceres, G.; Chhay, L.; O’Hara-Wild, M.; Petropoulos, F.; Razbash, S.; Wang, E.; Yasmeeen, F. *Forecast: Forecasting Functions for Time Series and Linear Models*; R Package Version 8.16, 2018. Available online: <https://cran.r-project.org/web/packages/forecast/forecast.pdf> (accessed on 2 January 2022).
26. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018. Available online: <https://www.R-project.org/> (accessed on 2 January 2022).