International Journal of Forecasting xxx (xxxx) xxx

Contents lists available at ScienceDirect

International Journal of Forecasting

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

A combination-based forecasting method for the M4-competition

Srihari Jaganathan [a,*], P.K.S. Prakash [b]

[a] UCB Inc., 1950 Lake Park Drive, Smyrna, GA 30080, USA
[b] ZS Associates, Safina Towers-South Block, 5th Floor, Ali Asker Road, Bengaluru 560052, Karnataka, India

ARTICLE INFO

Keywords:
M competition
Forecasting competitions
Combining forecasts
Time series forecasting
M4

ABSTRACT

Several researchers (Armstrong, 2001; Clemen, 1989; Makridakis and Winkler, 1983) have shown empirically that combination-based forecasting methods are very effective in real world settings. This paper discusses a combination-based forecasting approach that was used successfully in the M4 competition. The proposed approach was evaluated on a set of 100K time series across multiple domain areas with varied frequencies. The point forecasts submitted finished fourth based on the overall weighted average (OWA) error measure and second based on the symmetric mean absolute percent error (sMAPE).

1. Introduction

The five-month M4 forecasting competition that ended May 31, 2018, was an extension of three previous M-Competitions that began in 1982 (Makridakis et al., 1982, 1993; Makridakis & Hibon, 2000). The M4 competition data consisted of series from multiple business areas and of different frequencies, as is summarized on the M4 website.[1]

In literature several researchers (Armstrong, 2001; Clemen, 1989; Makridakis & Winkler, 1983) have empirically shown effectiveness of combination based forecast. Our motivation for participating in the M4 competition was to evaluate the effect, if any, of combination-based methods on the accuracy of forecasting when using both statistical and machine learning-based approaches. We proposed two types of combination-based forecasting approaches: evidence-based and optimization-based, and demonstrated in the M4 competition that simple combinations of forecasting models performed competitively well on 100 K univariate series.

The 24 forecasting methods that we used comprised mostly statistical models such as exponential smoothing, autoregressive integrated moving average (ARIMA), theta, Temporal HIErarchical Forecasting (THIEF), multiple aggregation prediction algorithm (MAPA), etc., as well as several machine learning models such as bagging and neural network models. Individual models were combined to produce ensemble forecasts based on the mean, trimmed mean, and median operators. The competition dataset was divided into training and validation data. Model selection was conducted based on the models' performances on a validation dataset. For yearly, quarterly, monthly, weekly, and hourly data, we chose a combination forecast based on a median operator because of its performance on the symmetric mean absolute percent error (sMAPE). For the daily frequency, we selected models that use a mean absolute scaled error (MASE) value that optimizes sMAPE.

The rest of this paper is organized as follows. Section 2 presents the principles that were used to guide the proposed combination-based methods, while Section 3 briefly describes the models that were used in the M4 competition. Section 4 details the mathematical formulations for the combination-based methods. Section 5 contains a discussion of model selection and the combination of forecasts, and is followed in Section 6 with our conclusions and proposed next steps.

* Corresponding author.
E-mail addresses: sriharitn@gmail.com (S. Jaganathan), prakash2@uwalumni.com (P.K.S. Prakash).

[1] https://www.m4.unic.ac.cy/the-dataset/.

https://doi.org/10.1016/j.ijforecast.2019.03.030
0169-2070/

2. Guiding principles for combination-based methods

This submission was influenced heavily by Armstrong's *Principles of Forecasting: A Handbook for Researchers and Practitioners* (Armstrong, 2001). Some of the guiding principles that we used are:

- **Combining forecasts**: Combinations or ensembles of forecasts with equal weights have been shown to perform well (Armstrong, 2001). With that in mind, we decided to combine the forecasts from individual models, weighting them equally, and to use mean or median operators because we did not know which methods performed best for a given time series.
- **Evidence-based methods**: With the exception of neural networks, almost all of our methods have been tested in earlier competitions such as the M1 or M3 forecasting competitions or benchmark datasets, and have been shown empirically to perform well.
- **Use appropriate methods for different frequencies:** Not all methods perform well for all frequencies, and in some instances the software package does not allow high frequency data, so we decided to use diverse forecasting methods for different frequencies.
- **Be conservative:** Research (Armstrong, Green, & Graefe, 2015) has shown that being conservative helps to improve the predictive accuracy in uncertain situations. We did not know the underlying data generating process for the time series in the M4 dataset, and as a consequence, we decided to use naïve and dampened methods as part of the ensemble.

We evaluated 24 methods in the M4 competition. Please note that we did not use all methods for all series. We selected our methods based on:

1. Guiding principles, as listed above.
2. The computational effort required; for example, monthly forecasts have large numbers of series, and therefore we avoided neural network models such as the long short-term memory (LSTM) because of their high computational requirements.
3. Their performances on hold-out sample data.

3. Methods

We used numerous statistical and machine learning methods in the M4 competition. Our key criterion for choosing a method was that it had been analyzed empirically using forecasting competition datasets or benchmarks. All of the methods used the default parameter settings in the appropriate software packages.

- **Naïve/snaïve:** We used naïve forecasts for a yearly frequency and seasonal naïve (snaïve) forecasts for other frequencies. The naïve method has been shown to be more accurate than nine of the 15 sophisticated methods on the original M competition data (Armstrong, 1984). This method was used to produce conservative forecasts according to the criteria of guiding principles listed earlier.
- **Exponential smoothing (ETS):** Automatic exponential smoothing methods are based on innovation state space models, as implemented in the R forecast package (Hyndman & Khandakar, 2008). Exponential smoothing has been used as a standard benchmark for comparing forecasting methods, and has been tested repeatedly on M3 competition data, on which it has performed well competitively (Hyndman, Koehler, Snyder, & Grose, 2002).
- **Dampened ETS:** A dampened version of ETS models. Dampened trend exponential methods have proven hard to beat in empirical forecasting competitions (Gardner & McKenzie, 2011). This method was used to produce forecasts that were conservative according to the guiding principles of the previous section.
- **Bagged ETS:** Bagged ETS is an approach that was proposed recently by Bergmeir, Hyndman, and Benítez (2016). This technique involves four steps:

  1. Use seasonal and trend decomposition using loess (STL) to decompose the time series into trend, seasonal, and remainder components.
  2. Bootstrap ($n = 100$ times, including the original series) the remainder component using a moving block bootstrap.
  3. Combine the trend and seasonal components and add back the bootstrapped series to the original series. This will result in a bootstrapped time series n.
  4. Generate the forecasts for a time series n and combine them by using either a mean or a median operator to generate a final forecast.

  Bagged ETS was used only on the monthly-frequency data, for which there was evidence in the M3 competition of this method's effectiveness.
- **Exponential smoothing (ES)/complex exponential smoothing (CES) and general exponential smoothing (GES)** (Svetunkov, 2018)**:**

  1. ES is an alternative to ETS models that allows series with frequencies > 24 to be used to combine forecasts from ETS models.
  2. CES was proposed by Svetunkov (2018) based on a new notion of "information potential", an unobserved time series component, and the theory of functions of complex variables.
  3. GES is a general exponential smoothing model.

  We used ES/CES/GES in a combination approach instead of ETS because these allow the modeling of frequencies > 24, whereas ETS only allows data with frequencies ≤ 24.
- **Multi-aggregation prediction algorithm (MAPA):** The MAPA (Kourentzes, Petropoulos, & Trapero, 2014) forecasting method involves three steps:

  1. Aggregation, which involves temporally aggregating the original time series by different groups of values.
  2. Forecasting the aggregated data from step 1 individually by using ETS and extracting forecasts of the level, trend, and seasonal components of the ETS for each series.

Please cite this article as: S. Jaganathan and P.K.S. Prakash, A combination-based forecasting method for the M4-competition. International Journal of Forecasting (2019), https://doi.org/10.1016/j.ijforecast.2019.03.030.

3. Combining the component forecasts generated in step 2 by using the mean or median operators.

Kourentzes et al. (2014) tested MAPA on the M3 competition data and other benchmarks and found that it outperformed the ETS method; however, a combination of MAPA and ETS improved the accuracy of the M3 competition forecasts significantly. Thus, MAPA and ETS are always together in our evidence-based ensemble framework.

- **Temporal hierarchical forecasting (THIEF):** Forecasting with THIEF (Athanasopoulos, Hyndman, Kourentzes, & Petropoulos, 2017) involves three steps:

  1. Use non-overlapping aggregation to aggregate the time series temporally up to the annual level, to produce temporal hierarchies.
  2. Generate forecasts for each aggregated series independently.
  3. Combine these optimally to produce forecasts that are reconciled across the short-, medium-, and long-term forecast horizons.

  Athanasopoulos et al. (2017) used M3 competition data to analyze the THIEF method extensively, and showed that it produced highly accurate forecasts in both the monthly and quarterly series.
- **Autoregressive integrated moving average (ARIMA):** Automatic ARIMA modeling is based on Hyndman and Khandakar's algorithm, as implemented in the forecast package in R (Hyndman & Khandakar, 2008). This method is the standard benchmark in any forecasting competition and has performed relatively well; hence, it was used as one of the methods for generating ensemble forecasts.
- **Theta:** The theta (Assimakopoulos & Nikolopoulos, 2000) method begins by seasonally adjusting data that are identified as seasonal. The resulting seasonally-adjusted data are then decomposed further into theta lines in order to capture the long- and short-term behaviors of time series data. The theta lines are forecast individually based on linear trends and simple exponential smoothing, then averaged. The data are then reseasonalized to generate the final forecast. The theta forecasting method was one of the most accurate methods used in the M3 forecasting competition.
- **Hybrid Theta:** A hybrid theta (Spiliotis, Assimakopoulos, & Nikolopoulos, 2019) method improves on the original theta method mentioned above. Instead of linear trends, the hybrid method uses nonlinear trends along with smoothing time series data and shrinking seasonal factors. This method was also tested on the M3 competition data, with very promising results in terms of forecast accuracy.
- **Forecast Pro:** Forecast Pro (Business Forecast Systems, 2018) is a commercial business forecasting software package that implements a variety of univariate and multivariate forecasting methodologies. It includes an expert selection algorithm that can automatically analyze a given time series, select and optimize an appropriate forecasting model, and generate the forecasts. Forecast Pro was one of the most accurate methods in the M3 competition, and hence was used in our combination approach.
- **Seasonal and trend decomposition using loess (STL) forecast:** The STL method decomposes a time series by using a seasonal component, a trend component, and the remainder component. It forecasts the trend component by means of the ETS, ARIMA, or theta method, then uses the seasonal component to reseasonalize the forecast. STL forecasting is accomplished using the forecast package in R (Hyndman & Khandakar, 2008). The authors used three STL methods in the M4 competition:

  - STL-ETS
  - STL-ARIMA
  - STL-THETA.

  Theodosiou (2011) showed that decomposing the time series into its trend, seasonal, and remainder components and then forecasting the decomposed series and aggregating the forecasts performed relatively well based on the M1 and M3 competition datasets.
- **Trigonometric Box–Cox transform, ARMA errors, trend, and seasonal components (TBATS):** TBATS (De Livera, Hyndman, & Snyder, 2011) is an innovations state space modeling framework for the forecasting of complex seasonal time series such as those with multiple seasonal periods, high-frequency seasonality, non-integer seasonality, and dual-calendar effects. The TBATS framework incorporates Box–Cox transformations, Fourier representations with time-varying coefficients, and ARMA error correction. This framework was used only for forecasting high-frequency data such as the weekly, daily, and hourly frequencies. Benchmark datasets were used to analyze it empirically, and found that it produced highly effective forecasts (De Livera et al., 2011). Hence, it was used for generating forecasts in the combination approach.
- **Double seasonal Holt-Winters (DSHW):** The DSHW (Taylor, 2003) is a Holt-Winters exponential smoothing formulation that accommodates two seasonalities. This technique is used in hourly data that exhibit multiple seasonalities, and performed well in our benchmark tests. Hence, DSHW was used in the ensemble of methods for our combination-based forecasting approach.
- **Multilayer perceptron (MLP)/extreme learning machines (ELM):** The feed-forward MLP and ELM (Kourentzes, 2017) neural network methods were used only for hourly data. This is one of the methods that has not been analyzed empirically. Future research will conduct an empirical analysis of this method.

Section 4 discusses the detailed mathematical formulation that is necessary for combining these different models.

Please cite this article as: S. Jaganathan and P.K.S. Prakash, A combination-based forecasting method for the M4-competition. International Journal of Forecasting (2019), https://doi.org/10.1016/j.ijforecast.2019.03.030.

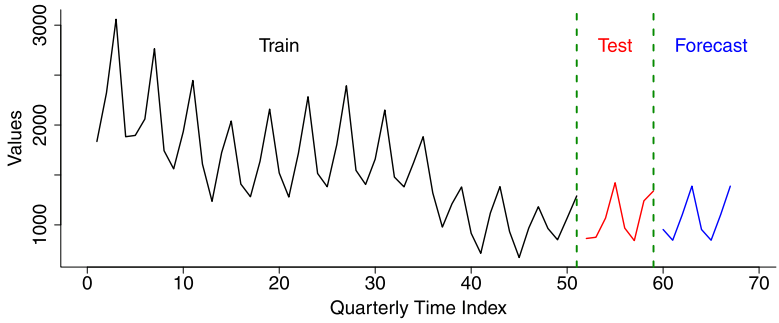

Fig. 1. Illustration of a quarterly series (Q5) data setup for the M4 competition.

4. Mathematical formulation for combining forecasts

This section presents the mathematical formulation that was used for combining the forecasts. The function estimate for univariate time series $Y = \{y(1), y(2), \ldots, y(n)\}$ is represented by $\widehat{F}_j(y(i))$ for series i and model j, where $i = 1, 2, \ldots, n$ forecasting series and $j = 1, 2, \ldots, m$ models. The M4 competition considered a weighted combination of two loss functions, namely the symmetric mean absolute percentage error (sMAPE) and the mean absolute scaled error (MASE) as proposed by Hyndman and Koehler (2006). The sMAPE for a series i and a model j can be evaluated as

$$sMAPE = \frac{1}{h}\sum_{t=1}^{h}\frac{2|y_t(i) - \widehat{F}_{jt}(y(i))|}{|y_t(i)| + |\widehat{F}_{jt}(y(i))|}, \tag{1}$$

where h is the forecasting horizon. The MASE for the ith series is evaluated as

$$MASE = \frac{1}{h}\frac{\sum_{t=1}^{h}|y_t(i) - \widehat{F}_{jt}(y(i))|}{\sum_{t=f+1}^{s}\frac{1}{s-f}|y_t(i) + y_{t-f}(i)|}, \tag{2}$$

where f represents the frequency of the data. For example, we used a frequency of 12 for a monthly time series frequency. The sMAPE helps in evaluating the forecasting accuracy, whereas MASE helps in understanding forecasting improvements with respect to the naïve model. A lower MASE represents a more accurate forecast.

We used two ensemble methods in the M4 competition:

1. An optimization-based ensemble that selects models for the ensemble based on MASE critical values, by optimizing the sMAPE on validation sets across series. This optimization-based ensemble approach aims to combine model selection and ensemble operations.
2. An evidence-based ensemble that relies on the guiding principles discussed in Section 2.

The optimization-based ensemble was used only for daily time series forecasts.

Let $\widehat{F^*}$ represent combination-based point forecasting that minimizes the expected loss function $\mathcal{L}(Y, \widehat{F^*}(Y))$. The loss function is defined as

$$\mathcal{L}(y(i), \widehat{F^*}) = \frac{1}{h}\sum_{t=1}^{h}\frac{2|y_t(i) - \widehat{F}_{jt}(y(i))|}{|y_t(i)| + |\widehat{F}_{jt}(y(i))|}. \tag{3}$$

$\widehat{F^*}(Y)$ is an ensemble of multiple models. For time t, the mean ensemble forecast is evaluated as:

$$\widehat{F^*}(y(i)) = \frac{1}{\sum_{j=1}^{m}\mathcal{I}(MASE_j)}\sum_{j=1}^{m}\widehat{F}_j(y(i))\mathcal{I}(MASE_j), \tag{4}$$

where $\mathcal{I}(MASE_j)$ is an indicator function based on a MASE value and is used to select a model for the ith time series ensemble. The MASE indicator function is represented by

$$\mathcal{I}(MASE_j) = \begin{cases} 1, & \text{if } MASE_j < C_k^* \\ 0, & \text{otherwise,} \end{cases} \tag{5}$$

where C_k^* is a critical value and $k \in \{hourly, daily, weekly, monthly, quarterly, yearly\}$. The indicator function $\mathcal{I}(MASE_j)$ defines an upper bound for the selection of a model. The model is selected for a time series if its MASE is less than a critical value. An evidence-based ensemble approach is a special case of an optimization-based ensemble in which the critical value C_k^* is set to infinity, and thus, all models were used. In contrast, in the optimization-based ensemble method, C_k^* is optimized and the upper bound of the selected method is not infinity; thus, the critical value acts as a selection criterion for a model to be considered for the ensemble.

The approach proposed in M4 used mainly the evidence-based ensemble method, which is a simple median of individual forecasts except for a daily time series that uses the optimization-based ensemble method. For the daily time series, the value of C_k^* is optimized by using a grid-search to minimize the loss function $\mathcal{L}(Y, \widehat{F^*}(Y))$. Also, we ensure that a forecast exists for every series by defining default models that will be chosen as an evidence-based ensemble if none of the series selected satisfies the C_k^* critical value requirement.

The optimization is performed by splitting the dataset into training and test (validation) sets. The testing horizon for each series is selected as its respective forecasting horizon. Fig. 1 illustrates an example of the data setup.

The pipeline is customized based on hourly, daily, weekly, monthly, quarterly, and yearly time series resolutions. For example, Fig. 1 shows test horizons for eight quarters of a quarterly series, which is the same as the forecasting horizon. Each model is evaluated independently, and a combination is selected based on optimal validation dataset performances.

Please cite this article as: S. Jaganathan and P.K.S. Prakash, A combination-based forecasting method for the M4-competition. International Journal of Forecasting (2019), https://doi.org/10.1016/j.ijforecast.2019.03.030.

Table 1
Model configurations at different frequencies.

Model	Hourly	Daily	Weekly	Monthly	Quarterly	Yearly
ETS	–	✓	–	✓	✓	✓
TBATS	✓	✓	✓	–	–	–
ARIMA	–	✓	–	✓	✓	✓
Damped ETS	–	–	–	✓	✓	✓
Naïve/SNaïve	–	–	–	✓	✓	✓
MAPA	–	–	–	✓	✓	✓
Theta	–	–	✓	✓	✓	✓
Hybrid theta	–	–	–	✓	✓	✓
Forecast Pro®	–	–	–	✓	✓	✓
STL-ETS	–	–	–	✓	✓	–
STL-ARIMA	–	–	–	✓	✓	–
STL-THETA	–	–	✓	✓	✓	–
THIEF ETS	–	–	–	✓	✓	–
THIEF ARIMA	✓	–	–	✓	✓	–
Bagged ETS	–	–	–	✓	–	–
ES	–	–	✓	–	–	–
CES	–	–	✓	–	–	–
THIEF THETA	–	–	✓	–	–	–
THIEF ES	–	–	✓	–	–	–
GES	✓	–	–	–	–	–
THIEF NAÏVE	✓	–	–	–	–	–
ELM	✓	–	–	–	–	–
MLP	✓	–	–	–	–	–
DSHW	✓	–	–	–	–	–

5. Model selection and ensemble of forecasts

Table 1 lists the model configurations used for different frequencies. The primary factors in our subjective judgment of model selection were:

1. The suitability of methods for a given frequency and prior evidence of their performance in previous competitions and benchmarks.
2. The availability of software.
3. Their performances in hold-out sample forecasts in the M4 dataset.

We took two approaches to the combination of individual forecasts in order to produce ensemble forecasts.

5.1. Optimization-based ensemble for daily time series

The daily series setup consists of three models: ETS, TBATS and ARIMA. ETS and TBATS were used as base models, and sMAPE was used as the criterion for optimizing ARIMA for a critical value of MASE. We determined the optimal critical value C_k^* (where k = daily) as in Eq. (5) by using a random sample of 1000 series out of 4227 daily time series to minimize the computational effort required. The daily time series was evaluated at multiple MASE critical values, ranging from (0, 4] with a resolution of 0.25. The output is shown in Table 2.

Table 2 shows that at the 0.25 MASE critical value, only 0.4% of the 1000 randomly-sampled time series used an ensemble of [ETS, TBATS, ARIMA]; the remainder of the time series used only the base model ([ETS, TBATS]) in the ensemble. With a high MASE critical value of 4.0, on the other hand, 81.8% of the daily series used [ETS, TBATS, ARIMA] in the ensemble and the remainder used only the base model. Fig. 2 illustrates the relationship between sMAPE and the MASE critical value.

Table 2
Average sMAPE values observed with different MASE critical values for an optimization-based ensemble of 1000 randomly-selected series.

Series no.	MASE critical value	Fraction using ARIMA[a]	sMAPE	MASE
1	0.25	0.4%	2.544	3.0659
2	0.5	1.9%	2.543	3.0658
3	0.75	7.1%	2.544	3.0660
4	1	16.4%	2.541	3.0641
5	1.25	25.2%	2.539	3.0620
6	1.5	35.0%	2.538	3.0615
7	1.75	43.1%	2.539	3.0619
8	2	48.7%	2.540	3.0624
9	2.25	53.5%	2.538	3.0594
10	2.5	56.9%	2.540	3.0604
11	2.75	61.2%	2.541	3.0628
12	3	64.9%	2.542	3.0631
13	3.25	69.5%	2.540	3.0615
14	3.5	74.7%	2.540	3.0624
15	3.75	78.3%	2.544	3.0690
16	4	81.8%	2.544	3.0698

[a]Fraction of series that use an ensemble of [ETS, TBATS, ARIMA].

Fig. 2 shows that sMAPE and the critical value of MASE follow a second-order relationship, and thus, it is important to find an optimal critical value (C_k^*). As can be seen in the figure, sMAPE is lowest when the critical MASE value is ∼2.0, and therefore we selected C_k^* for MASE as 2.0 for the ARIMA model. Thus, ARIMA is rejected in those series in which the MASE for ARIMA exceeds 2.0, and only the base models are used in the ensemble. Table 3 displays the performances of the optimized ensemble, the base model and the full ensemble. Clearly, the bounding of MASE has helped to decrease the sMAPE further.

We used sMAPE instead of OWA for the optimization setup. However, if we had tested the framework with OWA as an objective function, we could have gained improved accuracy. It is important to note that this approach was used only for the daily series; for all other series, we used a simple approach of combining individual models based on a median operator.

5.2. Evidence-based ensemble for non-daily time series

For all non-daily time series, we generated ensemble forecasts based on the evidence-based approaches outlined in the procedures for combining forecasts (Armstrong, 2001):

1. Use different methods: As Table 1 outlines, we generated individual forecasts using diverse methods that were appropriate to the frequencies.
2. Use at least five forecasts when possible: We generated more than five forecasts for all non-daily frequencies.
3. Use formal procedures to combine forecasts: We used simple combinations of forecasts with equal weights based on the trimmed mean, median, and mean of individual forecasts to generate ensemble forecasts. Based on the performances on the hold-out sample, we decided to use a median operator to generate ensemble forecasts.

Please cite this article as: S. Jaganathan and P.K.S. Prakash, A combination-based forecasting method for the M4-competition. International Journal of Forecasting (2019), https://doi.org/10.1016/j.ijforecast.2019.03.030.

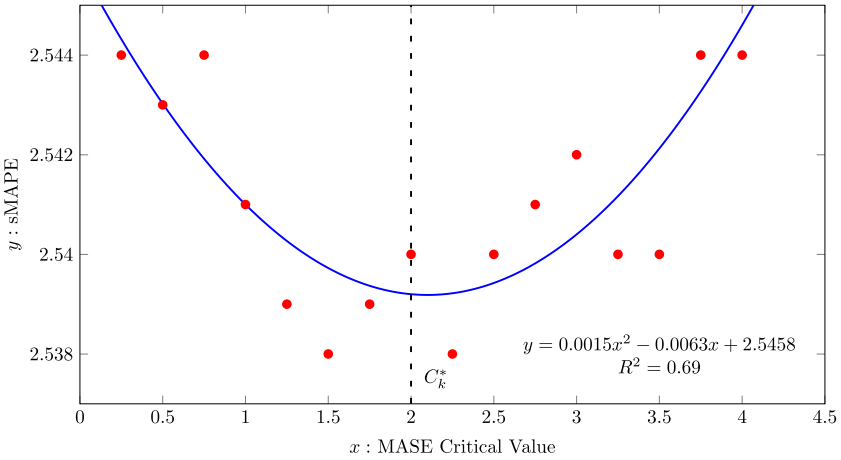

Fig. 2. sMAPE performances for daily time series with respect to MASE critical values.

Table 3
Comparison between the base model, the full ensemble and the optimized ensemble for all 4227 daily series.

Type	Model(s)	Critical value	sMAPE	MASE
Base model	[ETS, TBATS]	$[\infty, \infty]$	2.6339	3.8286
Full ensemble	[ETS, TBATS, ARIMA]	$[\infty, \infty, \infty]$	2.5622	3.0812
Optimized ensemble	[ETS, TBATS, ARIMA]	$[\infty, \infty, 2]$	**2.5396**	3.0624

Table 4
Proportions of series using the optimization- vs. evidence-based ensembles in the M4 competition.

Type	Frequency	No. of series	Proportion
Optimization-based ensemble	Daily	4,227	4.2%
Evidence-based ensemble	Non-daily series	95,773	95.8%
Total	All	100,000	100%

We experimented with two ensemble approaches for the daily and non-daily time series. Table 4 shows the proportions of the M4 competition series that used optimization-based (4.2%) vs. evidence-based (95.8%) ensembles. We believe strongly that the simplicity of evidence-based combination approaches means that they provide a simple and effective way of generating ensemble forecasts. We intend to consider this as a topic for future research, in which we plan to compare simple combinations with optimized weights to ensemble forecasts. The model's performance on the final holdout data of the M4 competition with respect to sMAPE and OWA was presented in a paper that summarized the M4 results (Makridakis, Spiliotis, & Assimakopoulos, 2018). The overall improvement obtained over the M4 benchmarks in terms of sMAPE was 6.8%, and that in terms of OWA was 6.2%.

6. Summary and discussion

This article has discussed a combination-based approach that combines statistical and machine learning techniques and was used for univariate forecasting in the M4 competition. Our proposed approach placed fourth overall using OWA, and second using sMAPE. A statistical test showed that the top six methods in the M4 competition did not differ significantly. The main finding from the competition was that combination approaches outperform individual methods. During the competition, the authors were hit significantly by the computing power required for machine learning models such as neural networks and bagging. We also used sMAPE or MASE to evaluate and select models and did not use OWA for model selection. In future, we plan to report the results from using OWA as a model selection criterion and compare them with those of other entrants to the competition.

Uncertainty in the forecast is a critical consideration in a final model. One of the main advantages of a combination-based method is that it not only helps us to arrive at better point forecasts, but also helps with the provision of better confidence intervals. One approach to the evaluation of prediction intervals for a combination-based method is to use a weighted combination of error-variances in which the weights are evaluated based on the model performance. Because we focused on point forecasting during the M4 competition, the problem of extracting the right prediction interval from a forecast using a combination-based method is expected to form part of our future research.

Another critical dimension for future study of the combination-based method is the optimization of the computing effort. Depending on the methods used, the

Please cite this article as: S. Jaganathan and P.K.S. Prakash, A combination-based forecasting method for the M4-competition. International Journal of Forecasting (2019), https://doi.org/10.1016/j.ijforecast.2019.03.030.

combination-based method can become quite expensive computationally. For example, the bagging ETS took approximately 24 h for 48K monthly series with 192 cores running in parallel, and approximately seven hours for 4227 daily series. Similarly, the neural network method for 414 hourly forecasts took approximately eight hours with a 128-core machine. Future research involves selecting the right setup model, based on the local properties of a time series, to achieve an accuracy similar to our results but with a lower computational time.

Program for replication

All open source R programs used for the competition can be accessed in github: https://github.com/M4Competition/M4-methods/tree/forecaster18-patch-2/srihari_prakash.

Acknowledgments

The authors gratefully acknowledge feedback from the associate editor and reviewers that significantly helped to improve the quality of our paper.

References

Armstrong, J. S. (1984). Forecasting by extrapolation: Conclusions from 25 years of research. *Interfaces*, *14*(6), 52–66.
Armstrong, J. S. (2001). *Principles of forecasting: a handbook for researchers and practitioners, Vol. 30*. Springer Science & Business Media.
Armstrong, J., Green, K. C., & Graefe, A. (2015). Golden rule of forecasting: Be conservative. *Journal of Business Research*, *68*(8), 1717–1731.
Assimakopoulos, V., & Nikolopoulos, K. (2000). The theta model: a decomposition approach to forecasting. *International Journal of Forecasting*, *16*(4), 521–530.
Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., & Petropoulos, F. (2017). Forecasting with temporal hierarchies. *European Journal of Operational Research*, *262*(1), 60–74.
Bergmeir, C., Hyndman, R. J., & Benítez, J. M. (2016). Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation. *International Journal of Forecasting*, *32*(2), 303–312.
Business Forecast Systems, Inc., 2018, Forecast Pro TRAC. Version:5.1, https://www.forecastpro.com/.
Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*(4), 559–583.
De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, *106*(496), 1513–1527.
Gardner, E. S., & McKenzie, E. (2011). Why the damped trend works. *The Journal of the Operational Research Society*, *62*(6), 1177–1180.
Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, *26*(3), 1–22.
Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688.
Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, *18*(3), 439–454.
Kourentzes, N. (2017). nnfor: Time series forecasting with neural networks, R package version 0.9.2.
Kourentzes, N., Petropoulos, F., & Trapero, J. R. (2014). Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting*, *30*(2), 291–302.
Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., et al. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, *1*(2), 111–153.
Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K., et al. (1993). The M2-competition: A real-time judgmentally based forecasting study. *International Journal of Forecasting*, *9*(1), 5–22.
Makridakis, S., & Hibon, M. (2000). The M3-competition: results, conclusions and implications. *International Journal of Forecasting*, *16*(4), 451–476.
Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 competition: Results, findings, conclusions and way forward. *International Journal of Forecasting*, *34*(4), 802–808.
Makridakis, S., & Winkler, R. L. (1983). Averages of forecasts: Some empirical results. *Management Science*, *29*(9), 987–996.
Spiliotis, E., Assimakopoulos, V., & Nikolopoulos, K. (2019). Forecasting with a hybrid method utilizing data smoothing, a variation of the theta method and shrinkage of seasonal factors. *International Journal of Production Economics*, *209*, 92–102.
Svetunkov, I. (2018). smooth: Forecasting using state space models, R package version 2.4.4.
Taylor, J. W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing. *The Journal of the Operational Research Society*, *54*(8), 799–805.
Theodosiou, M. (2011). Forecasting monthly and quarterly time series using STL decomposition. *International Journal of Forecasting*, *27*(4), 1178–1195.