Improving forecasting by estimating time series structural components across multiple frequencies

CrossMark

Nikolaos Kourentzes [a,*], Fotios Petropoulos [a], Juan R. Trapero [b]

[a] *Lancaster University Management School, Department of Management Science, Lancaster, Lancashire, LA1 4YX, UK*
[b] *Universidad de Castilla-La Mancha, Departamento de Administracion de Empresas, Ciudad Real 13071, Spain*

A R T I C L E   I N F O

A B S T R A C T

Identifying the most appropriate time series model to achieve a good forecasting accuracy is a challenging task. We propose a novel algorithm that aims to mitigate the importance of model selection, while increasing the accuracy. Multiple time series are constructed from the original time series, using temporal aggregation. These derivative series highlight different aspects of the original data, as temporal aggregation helps in strengthening or attenuating the signals of different time series components. In each series, the appropriate exponential smoothing method is fitted and its respective time series components are forecast. Subsequently, the time series components from each aggregation level are combined, then used to construct the final forecast. This approach achieves a better estimation of the different time series components, through temporal aggregation, and reduces the importance of model selection through forecast combination. An empirical evaluation of the proposed framework demonstrates significant improvements in forecasting accuracy, especially for long-term forecasts.
© 2013 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

1. Introduction

In forecasting, the selection and parameterisation of models are of critical importance, as they are tied to accurate and reliable predictions. Modern organisations have to produce large numbers of forecasts. It is desirable to identify and parameterise the appropriate model automatically for each series, in order to reduce the associated workload. This endeavour has met with mixed success in both research and practice, resulting in De Gooijer and Hyndman (2006) identifying it as an open research question. The lack of statistical expertise in organisations complicates the problem further (Hughes, 2001), and a relatively simple solution would be desirable for practical implementations.

In this paper, we propose a framework that mitigates the issue of model selection, while improving the forecast-ing accuracy, by taking advantage of temporal aggregation and forecast combination.

It is possible to emphasise different time series characteristics by transforming the original data to alternative time frequencies. We propose to aggregate a time series into multiple lower frequencies, i.e., a monthly time series becomes bi-monthly, quarterly and so on. At each aggregation level, different features of the time series are highlighted and attenuated (Andrawis, Atiya, & El-Shishiny, 2011). At lower aggregations (high frequency time series), periodic components such as seasonality will be prominent. As the aggregation level increases, high frequency signals are filtered out, and more importance is given to the lower frequency components, such as the level and trend of a time series. Intuitively, we expect to capture the seasonal elements of a time series better at lower aggregation levels (high frequency data). The opposite is true for the level and trend, which are highlighted at higher aggregation levels. Consequently, our motivation for such aggregation is to facilitate the identification, selection and parameter estimation of forecasting models.

---

* Corresponding author. Tel.: +44 1524 592911.
E-mail address: n.kourentzes@lancaster.ac.uk (N. Kourentzes).

After temporal aggregation, we produce a forecast at each aggregation level, using the appropriate forecasting method. For this purpose we use the exponential smoothing family of methods, which provides a holistic framework for modelling all archetypal types of time series, whether level, trended or seasonal. The calculation of numerous forecasts at different frequencies inevitably leads to the use of combination schemes. The combination step offers further improvements in the robustness and accuracy of the resulting forecast.

Instead of combining the forecasts from the different aggregation levels directly, we propose to first break down the forecasts to their time series components, then combine these. This is motivated by potential differences in the ways in which the components appear or are transformed at different temporal aggregation levels. For example, if a seasonal monthly time series is aggregated into an annual one, it will have no seasonality. Simply combining monthly and annual forecasts would halve the seasonal signal. Therefore, the combination of the seasonal information must only be done for the aggregation levels at which it exists, whilst for the level and trend, the combination can be performed at any level. Moreover, combining forecasts instead of time series components does not allow us to discriminate between the desirable elements of each aggregation level that is to be combined. Further advantages of combining components instead of forecasts are discussed in the description of the proposed algorithm. Exponential smoothing provides such component information directly (Gardner, 2006). The combined components are subsequently joined in order to produce forecasts of the original time series.

The key advantages of the proposed framework are: (i) it lessens the importance of model selection, a principal problem in time series modelling; (ii) it takes advantage of gains in forecasting accuracy from both temporal aggregation and forecast combination, which have been demonstrated separately in the literature; (iii) it makes use of exponential smoothing, a widely established and well researched forecasting method that is implemented in several forecasting support systems; and (iv) it is simple, thus allowing transparency and direct implementation in organisations.

We conduct a large empirical evaluation of the proposed framework against exponential smoothing. We also use the direct combination of forecasts produced at different aggregation levels as a benchmark, in order to demonstrate the benefits of considering the time series components separately. We test the robustness of the proposed *Multi Aggregation Prediction Algorithm* (MAPA) by evaluating its performance, in terms of accuracy and forecasting bias, across datasets with different sampling frequencies. We find that MAPA has a promising forecasting performance, with major improvements for long-term forecasts.

The rest of the paper is organized as follows: in the next section we discuss the benefits of temporal aggregation and model combination in the literature, providing further motivation for our work. Section 3 describes the proposed MAPA, while Section 4 outlines the experimental setup and presents the evaluation results, followed by a discussion in Section 5. A further refinement of the model is investigated in Section 6, and concluding remarks are given in Section 7.

2. Background research

2.1. Aggregation and forecasting

There are two types of aggregation in the forecasting literature (Babai, Ali, & Nikolopoulos, 2012). On the one hand, temporal aggregation refers to time (non-overlapping) aggregation for a specific time series. This converts a high frequency time series into a lower frequency time series, according to the selected aggregation level. For example, we can derive the annual demand of a monthly series by setting the aggregation level equal to 12 periods. On the other hand, cross-sectional or hierarchical aggregation refers to the demand aggregation (bottom-up) of multiple items or stock keeping units (SKUs) for the formation of families of products.

Temporal aggregation has been investigated in the context of both ARIMA models and GARCH processes. For a thorough overview of the effects of temporal aggregation on time series models, the reader is encouraged to refer to the study by Silvestrini and Veredas (2008). Even though it has no theoretical support (Wei, 1990), aggregation has been shown empirically to work remarkably well for ARIMA models, in terms of both forecasting accuracy (for example, see Abraham, 1982; Amemiya & Wu, 1972) and bias reduction (Mohammadipour & Boylan, 2012; Souza & Smith, 2004).

The consequences of temporal aggregation and systematic sampling, in terms of lag lengths (Brewer, 1973) and dynamic relationships between variables (Weiss, 1984), have also been investigated, giving useful theoretical insights into the stability of ARMA and ARMAX models. Moreover, Drost and Nijman (1993) showed that, in the case of univariate GARCH models, the variance parameters of the low frequency model generally depend on the mean, variance, and kurtosis parameters of the high frequency model. Hafner (2008) derived results for the temporal aggregation of multivariate GARCH(1,1) processes, and concluded that the dynamics of the aggregated processes can be acquired in a very simple way. As a result, plenty of empirical and theoretical work has been done on ARIMA models and GARCH processes in regard to temporal aggregation. On the other hand, the literature on the effects of temporal aggregation with exponential smoothing models is much more limited, even though the method itself is applied widely in practice (Gardner, 2006).

Spithourakis, Petropoulos, Babai, Nikolopoulos, and Assimakopoulos (2011) examined the efficiency of a temporal aggregation framework for widely used forecasting techniques. The results, based on the monthly data from the M3-Competition (Makridakis & Hibon, 2000), demonstrated significantly improved forecasting accuracies for Naïve, SES and Theta (Assimakopoulos & Nikolopoulos, 2000). Athanasopoulos, Hyndman, Song, and Wu (2011) showed that, in the tourism industry, aggregated forecasts (produced from high frequency data) were more accurate than forecasts produced from low frequency data directly. They were the first to investigate the impact of temporal aggregation on the performance of the exponential smoothing family of methods empirically, though their study was limited to aggregating monthly and quarterly data to yearly. Their forecasting evaluation was performed

only at the lowest frequency (yearly data). However, the effects of temporal aggregation on exponential smoothing models remain largely unexplored, particularly across different time frequencies.

Research has also focused on the application of temporal aggregation to count and intermittent data. In these cases, aggregation is intuitively appealing, as it is likely that the resulting time series will have a lower degree of intermittency. Willemain, Smart, Shockor, and DeSautels (1994) were the first to examine the impacts of temporal aggregation empirically in the context of intermittent demand. Although it was limited to just 16 series, their research demonstrated significant improvements in the behaviour of Croston's method when the data were aggregated in weekly time buckets. Nikolopoulos, Syntetos, Boylan, Petropoulos, and Assimakopoulos (2011) performed a more extensive empirical evaluation of the effects of the aggregation level in the performances of the random walk and the Syntetos and Boylan approximation of Croston's method (Syntetos & Boylan, 2005). They proposed an Aggregate-Disaggregate Intermittent Demand Approach (ADIDA) algorithm that acts as a "self improving mechanism" by reducing the bias and improving accuracy. Babai et al. (2012) showed that ADIDA also results in improved customer service levels. Spithourakis, Petropoulos, Nikolopoulos, and Assimakopoulos (forthcoming) presented a systemic view of the ADIDA framework. Using simulated examples, they demonstrated that there is a reduction in the output noise as the aggregation level increases. They concluded that temporal aggregation up to a limited level is considered beneficial, as the randomness is smoothed and the long-term features of the series are enhanced. At the same time, they found that large time buckets suffer from excess smoothing in trending or seasonal data.

One of the main issues is the selection of the optimum temporal aggregation level for a set of series, or for each series individually, as there is no evidence that this level is universal across a dataset. In this direction, Nikolopoulos et al. (2011) proposed setting the aggregation level of each series to be the lead time plus one review period. This heuristic demonstrated a good forecasting accuracy. In order to identify the optimal aggregation level for each series empirically, Spithourakis et al. (2011) used various in-sample criteria and found significant improvements for aggregation levels of up to around 12 periods.

The proposed MAPA builds on these findings. However, it does not focus on a single pre-determined level of aggregation, due to the absence of any theoretical understanding of the best way to choose one. While ADIDA uses a single aggregation level, MAPA considers a time series under multiple temporal aggregation levels simultaneously, with the aim of capitalising on the advantages of each aggregated and original time series. Since no single level is selected, different forecasts are produced for the same item. The resulting information can be combined in a single prediction, the output of MAPA.

2.2. Combining forecasts for different aggregation levels

Combining has been considered widely to be beneficial for forecasting (Bates & Granger, 1969; Clemen, 1989; Makridakis & Winkler, 1983). Averaging statistical approaches leads to improvements in predictive accuracy. Combining also results in lower levels of uncertainty, by reducing the variance of the forecasting errors (Hibon & Evgeniou, 2005). Many recent studies have examined sophisticated weight selection processes for combining forecasts (He & Xu, 2005; Kolassa, 2011; Taylor, 2008); however, simple approaches are considered to be robust and perform reasonably well relative to more complex methods (Clemen, 1989; Hibon & Evgeniou, 2005; Jose & Winkler, 2008; Timmermann, 2006). According to Kolassa (2011), there is little consensus in the literature as to whether the mean or median of forecasts should be selected, in terms of forecasting performances.

The combination of forecasts produced from different frequencies of the same data is intuitively appealing, with the aim of capturing different dynamics of the data at hand. To the best of our knowledge, Cholette (1982) was the first to investigate the advantages of this scheme, by modifying the monthly forecasts produced from an ARIMA model through the integration of judgmental forecasts derived from experts using data of alternative frequencies. Trabelsi and Hillmer (1989) investigated the problem of combining forecasts constructed on series of different frequencies, again using ARIMA models. A process used to specify a high frequency model for multiple series sampled at different frequencies is discussed by Casals, Jerez, and Sotoca (2009). They demonstrated theoretically that aggregation may have a positive effect on the predictive accuracy, by combining forecasts from data of different frequencies efficiently. Andrawis et al. (2011) investigated the benefits of combining short- and long-term forecasts. They used monthly and yearly time series and investigated the best method for combining the forecasts. In order to make the combination feasible, they first deseasonalised the monthly time series. They concluded that the combination can be beneficial for forecasting accuracy. Granger (1993) argued that temporal aggregation should be done with care, especially in the presence of cointegration, as it may have a variety of effects and implications.

3. Forecasting through temporal aggregation of time series components

The proposed framework can be separated into three steps: (i) the aggregation step; (ii) the forecasting step; and (iii) the combination step. We will discuss each step individually, then present MAPA holistically. A visual overview of MAPA is presented in Fig. 1, outlining the three different steps. We will refer to this figure in the detailed description of the algorithm.

3.1. Time series temporal aggregation

Let Y be a time series, with observations y_t and $t = 1, \ldots, n$, sampled at a given frequency. We can aggregate this time series by considering consecutive groups of values of the original time series in sets of length k. We denote the temporally aggregated time series as $Y^{[k]}$, with
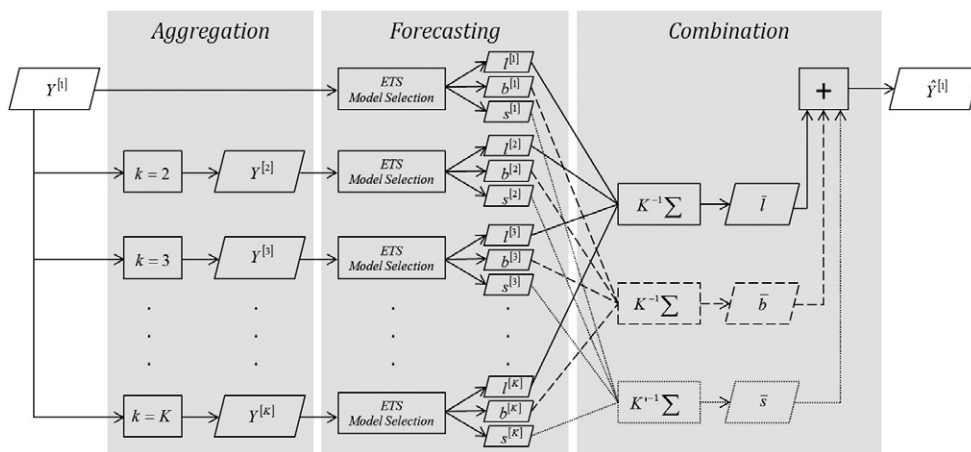
Fig. 1. Flowchart of the proposed Multi Aggregation Prediction Algorithm (MAPA).

observations $y_i^{[k]}$ and $i = 1, 2, \ldots, n/k$, such as:

$$y_i^{[k]} = k^{-1} \sum_{t=1+(i-1)k}^{ik} y_t. \qquad (1)$$

Obviously, for $k = 1, y_i^{[1]} = y_t$. Hereafter, for consistency, we will only use $y_i^{[k]}$. Observe that we consider the arithmetic mean as the aggregation operator. We do this in order to keep the level of the aggregated time series $Y^{[k]}$ similar to the level and units of the original time series $Y^{[1]}$. The resulting $Y^{[k]}$ has $\lfloor n/k \rfloor$ observations. If the remainder of the division n/k is not zero for a specific aggregation level k, then some observations will not be taken into account when forming the temporal aggregation buckets. In this case, we remove $n - \lfloor n/k \rfloor$ observations from the beginning of the time series in order to be able to form complete aggregation buckets. In theory, we can aggregate a time series for all $k = 2, \ldots, n$, but in practice it is preferable for $k \ll n$ in order to leave an adequate sample for parameterising a forecasting method on $Y^{[k]}$.

An example is provided in Fig. 2, where a monthly time series is aggregated for $k = 1, 3, 5, 12$. Plots of the time series are provided, together with the corresponding power spectra. For $k = 1$, a clear seasonal signal is evident in the power spectrum. The same is true for $k = 3$, although the variance it explains is now smaller, and a shift in frequency is also clear. For higher aggregation levels, the seasonal component cannot be observed in either the time series or power spectra plots; in fact, the level and trend components dominate.

In the first step of MAPA, as shown in Fig. 1, K time series are produced through time series temporal aggregation.

3.2. Forecasting

After constructing the aggregate $Y^{[k]}$ series, we fit appropriate forecasting models to each one. It is desirable to consider the different time series components that appear at each aggregation level, instead of the forecasts, so as to distinguish the aggregation levels at which the seasonal component can appear. Furthermore, the focus

on the components simplifies the forecast combination, as a combination of different level components or different seasonal components can be done directly, with no special pre-processing. On the other hand, combining level, trended and seasonal forecasts is not trivial. For example, combining a seasonal and a non-seasonal forecast would result in the seasonal signal being halved. Previous studies have tackled this issue by seasonally adjusting the original data (for example, Andrawis et al., 2011), and thus forcing a deterministic seasonality, which may be inappropriate, as well as requiring it to be characterised as either additive or multiplicative. This shift, from combining forecasts to combining forecasted time series components, is one of the key innovations of the proposed algorithm.

The exponential smoothing forecasting method lends itself to this purpose, as one can extract the level, trend and seasonal components of the fitted model directly. In this paper, we will refer to it as ETS, from *ExponenTial Smoothing*, as was proposed by Hyndman, Koehler, Ord, and Snyder (2008). ETS is a family of methods, designed for forecasting time series with different characteristics, covering a wide range of additive and multiplicative combinations of level, trend, seasonality and noise (Gardner, 1985, 2006; Taylor, 2003). ETS has been shown to provide a relatively good forecasting accuracy over a large number of time series from many different sources (Hyndman et al., 2008; Makridakis & Hibon, 2000). The state space framework, proposed by Hyndman, Koehler, Snyder, and Grose (2002), offers a statistical rationale to support the ETS models. Under the state space formulation, each of the components of the time series is modelled in the state vector, and is available for further manipulations.

ETS models capture and smooth the existing structural components of a time series: level, trend and seasonality. The components of the method can interact in either an additive or a multiplicative way. The trend can be linear or damped. For a full description of ETS, together with a discussion on how to optimise its smoothing parameters, see Hyndman et al. (2008). For each time series, the appropriate components must be selected. Hyndman et al. (2002) propose that Akaike's Information Criterion be used for selecting between the different types of ETS
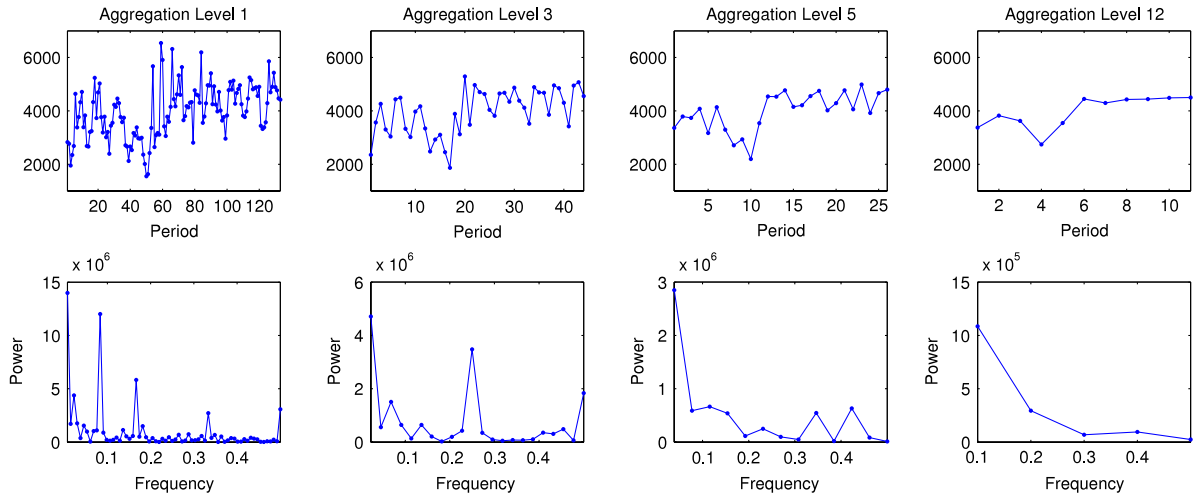
Fig. 2. Time series plots and power spectra for $k = 1, 3, 5, 12$.

automatically, with good accuracy results. Since our focus here is not on ETS, we follow their suggestion, which is widely regarded as being the state-of-the-art, on how to select the appropriate model.

For each temporally aggregated $Y^{[k]}$ time series, an ETS is fitted. We are interested in the last state vector $\boldsymbol{x}_i^{[k]}$ of ETS. The state vector contains the updated values for each time series component: level (l_i), trend (b_i) and seasonality (s_i), where available; $\boldsymbol{x}_i^k = (l_i, b_i, s_i, s_{i-1}, \ldots, s_{i-S+1})'$. Using this information and knowing the type of ETS fitted, we can produce forecasts for any forecast horizon desired. Note that, for $k > 1$, the desired forecast horizon h is adjusted for each aggregation level to $h^{[k]} = \lceil h/k \rceil$. Moreover, for the original time series, horizon $h^{[1]} = h$.

The seasonality may disappear and reappear at different aggregation levels. If $Y^{[1]}$ has a seasonal period of S, then seasonality will be considered again only if S/k is an integer number and $k < S$. For example, if the original periodicity is $S = 12$ for monthly data, then seasonality will be modelled again for k equal to 2, 3, 4 and 6. Potential non-integer seasonalities are not considered, due to the inability of ETS and other conventional forecasting methods to model them.

The numerical values of the additive and multiplicative components of $\boldsymbol{x}_i^{[k]}$ are not directly comparable. It is possible that an additive ETS could be fitted at one aggregation level and a multiplicative ETS at a different one. This can increase the complexity of the subsequent combination of the components from all $Y^{[k]}$. We propose that multiplicative components be translated to additive ones. Using the components in the state vector $\boldsymbol{x}_i^{[k]}$, we construct predictions for each level (l_i), trend (b_i) and seasonality (s_i), where available, translating to additive when needed, using the formulas in Table 1. The formulas can be derived trivially from the multi-step-ahead ETS forecast formulation, and are based on adding the level information to the multiplicative trend and seasonality. Note that parameter optimisation and the calculation of the updated states are done using the original formulation of ETS (Hyndman et al., 2008), independently at each aggregation level. Once

the components have been translated to additive ones, the model parameters at each level are no longer needed, and only the values of the translated components at each point t are used hereafter.

The additive translation of the components is only used for constructing the out-of-sample component predictions. Although the nature of the interaction of the error with the rest of the components, i.e., additive or multiplicative, is important for correctly optimising ETS and updating its state vector, it does not affect the out-of-sample prediction of the components, and Table 1 can be used in both cases.

Once all of the components have been translated to additive, we can proceed to the next step of the proposed algorithm. For notational consistency, we will add a superscript k to each component to indicate its respective level of aggregation. Note that under this additive reformulation, forecasted values of $Y^{[k]}$ can be calculated as a simple addition of the predicted components:

$$\hat{y}_{i+h^{[k]}}^{[k]} = l_{i+h^{[k]}}^{[k]} + b_{i+h^{[k]}}^{[k]} + s_{i-S+h^{[k]}}^{[k]}. \tag{2}$$

The forecasting of different aggregation levels, for a given time period, outputs different numbers of values. For example, if the original time series $Y^{[1]}$ is sampled at a monthly frequency, predicting one year ahead requires 12 forecasted values, while the $Y^{[12]}$ aggregate series requires only one value for the same forecast. It is desirable, in order to facilitate the combination of the forecasts, to have all aggregate forecasts on the time scale of the original time series $Y^{[1]}$. For any value $z_i^{[k]}$, this can be done using:

$$z_t = \sum_{j=1}^{k} \omega_j z_i^{[k]}, \tag{3}$$

with $t = 1, 2, \ldots, n$ and $i = \lceil t/k \rceil$. Essentially, Eq. (3) spreads the ith value to k observations in the original time domain t, acting as a piecewise constant interpolation.[1]

---

[1] More complex interpolation methods can also be used for this purpose; however, they introduce additional undesired complexity, particularly for aligning time across aggregation levels.

Table 1
Component prediction in the additive formulation.

Trend	Seasonal		
	N	A	M
N	$l_{i+h} = l_i$	$l_{i+h} = l_i$ $s_{i-S+h} = s_{i-S+h}$	$l_{i+h} = l_i$ $s_{i-S+h} = (s_{i-S+h} - 1)l_{i+h}$
A	$l_{i+h} = l_i$ $b_{i+h} = hb_i$	$l_{i+h} = l_i$ $b_{i+h} = hb_i$ $s_{i-S+h} = s_{i-S+h}$	$l_{i+h} = l_i$ $b_{i+h} = hb_i$ $s_{i-S+h} = (s_{i-S+h} - 1)(l_{i+h} + b_{i+h})$
Ad	$l_{i+h} = l_i$ $b_{i+h} = \sum_{j=1}^{h} \phi^j b_i$	$l_{i+h} = l_i$ $b_{i+h} = \sum_{j=1}^{h} \phi^j b_i$ $s_{i-S+h} = s_{i-S+h}$	$l_{i+h} = l_i$ $b_{i+h} = \sum_{j=1}^{h} \phi^j b_i$ $s_{i-S+h} = (s_{i-S+h} - 1)(l_{i+h} + b_{i+h})$
M	$l_{i+h} = l_i$ $b_{i+h} = (b_i^h - 1)l_{i+h}$	$l_{i+h} = l_i$ $b_{i+h} = (b_i^h - 1)l_{i+h}$ $s_{i-S+h} = s_{i-S+h}$	$l_{i+h} = l_i$ $b_{i+h} = (b_i^h - 1)l_{i+h}$ $s_{i-S+h} = (s_{i-S+h} - 1)(l_{i+h} + b_{i+h})$
Md	$l_{i+h} = l_i$ $b_{i+h} = (b_i \sum_{j=1}^{h} \phi^j - 1)l_{i+h}$	$l_{i+h} = l_i$ $b_{i+h} = (b_i \sum_{j=1}^{h} \phi^j - 1)l_{i+h}$ $s_{i-S+h} = s_{i-S+h}$	$l_{i+h} = l_i$ $b_{i+h} = (b_i \sum_{j=1}^{h} \phi^j - 1)l_{i+h}$ $s_{i-S+h} = (s_{i-S+h} - 1)(l_{i+h} + b_{i+h})$

This approach is fully equivalent to calculating point forecasts directly to the aggregated level and then reverting the forecasts back to the original level through disaggregation with equal weights ($\omega_j = 1/k$) (Nikolopoulos et al., 2011; Spithourakis et al., 2011).

We apply Eq. (3) to all $l_i^{[k]}$, $b_i^{[k]}$ and $s_i^{[k]}$, resulting in up to k different estimations for each of the time series components in the original time domain. Fig. 3 illustrates this for the estimated trend component of a time series. The trend component estimated at aggregation levels $k = (1, 4, 8)$ is displayed. Note that these have different slopes and numbers of observations, and they can be additive or multiplicative, damped or not. These are subsequently transformed to additive using Table 1 and back to the original time scale using Eq. (3).

In some cases, not all ETS components are fitted at all aggregation levels. For instance, ETS may capture the trend in some aggregation levels, but not in others. When there is no estimate of a particular component at an aggregation level, we set its predictions equal to zero. The rationale is that if a component is left out, then it is not identified in the time series, having zero impact on the forecast; in other words, it has zero additive contribution.

In the second step of the MAPA algorithm, as illustrated in Fig. 1, an ETS model is fitted for each aggregation level, and its components are extracted. These are subsequently transformed into additive ones and in the original time domain of $Y^{[1]}$. The output of this step is sets of level, trend and seasonal components from each aggregation level, ready to be combined in the next step.

3.3. Combined component forecast

In the third and final step of the MAPA algorithm, as shown in Fig. 1, the different components that were calculated before are combined to produce the final level, trend and season estimates, which are subsequently used to produce the final forecast.

Combining time series components instead of forecasts makes the process substantially simpler. We combine level, trend and seasonal components separately, which

are subsequently used to construct the final forecast. Since the aggregation of the time series was done using the arithmetic mean, we do not have to adjust the level of the resulting components further. In this work, we consider only two combination methods that have been shown to perform well and be robust: the unweighted mean and the median (see Agnew, 1985; Clemen, 1989; Clements & Hendry, 2007, chap. 1; Miller, Clemen, & Winkler, 1992; Palm & Zellner, 1992; Timmermann, 2006. A detailed comparison of the various different combination schemes is outside the scope of this paper.

In the case of the unweighted mean, the components are combined as follows:

$$\bar{l}_{t+h[1]} = K^{-1} \sum_{k=1}^{K} l_{t+h[1]}^{[k]}, \tag{4}$$

$$\bar{b}_{t+h[1]} = K^{-1} \sum_{k=1}^{K} b_{t+h[1]}^{[k]}, \tag{5}$$

$$\bar{s}_{t+h[1]} = K'^{-1} \sum_{k=1}^{K'} s_{t+h[1]}^{[k]}, \quad \text{if } (S/k) \in \mathbb{Z} \text{ and } k < S, \tag{6}$$

where K is the maximum aggregation level considered and K' is the number of aggregation levels where seasonality may be identified, i.e., when S/k results in an integer and $k < S$, as was discussed in Section 3.1.

The combination is illustrated in Fig. 4. Estimates of the level and trend components of a trended time series are provided for temporal aggregation levels $k = (1, 4, 8)$, as well as for the resulting combined components. Although in the examples we use only a few aggregation levels for illustrative purposes, we propose to build the combined MAPA components based on multiple continuous aggregation levels, resulting in smoother combined components.

Alternatively, we explore the use of the median instead of the mean in the construction of $\bar{l}_{t+h[1]}$, $\bar{b}_{t+h[1]}$ and $\bar{s}_{t+h[1]}$. The median is more robust to extreme values, and will therefore be affected less by poorly estimated components. If the ETS, applied at different aggregation levels, identifies different components, the median will favour the most
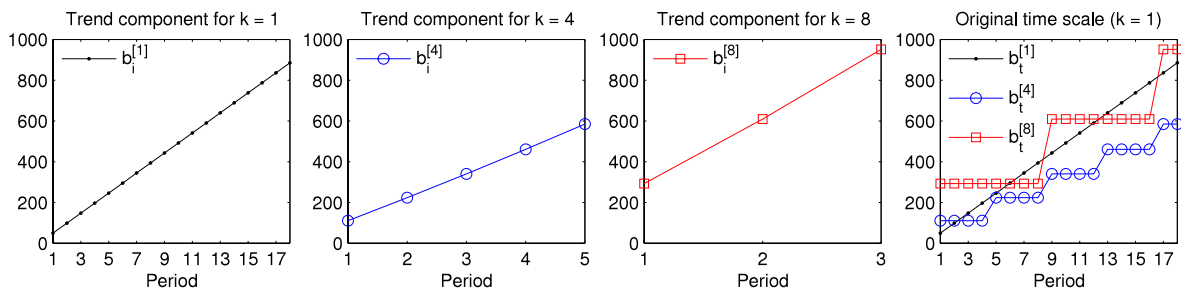
Fig. 3. Estimated trend component for $k = (1, 4, 8)$ and in the original time scale.
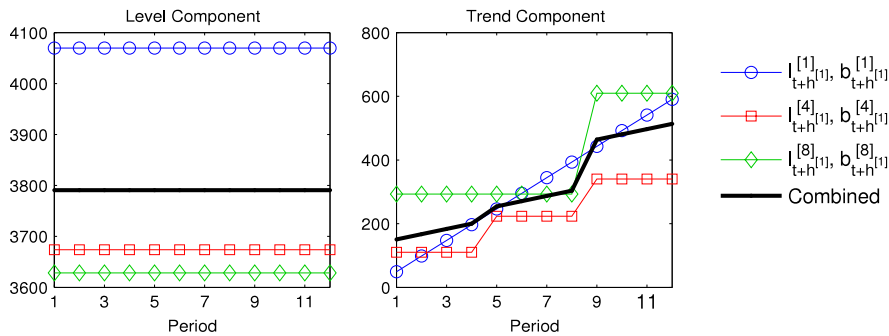

Fig. 4. Level and trend components for $k = (1, 4, 8)$ and the resulting combined $\bar{l}_{t+h^{[1]}}$ and $\bar{b}_{t+h^{[1]}}$ components.

popular component type. On the other hand, the mean considers all components equally, whether disagreeing or not.

To produce the final forecast for $h^{[1]}$ steps ahead, the forecast horizon of the original time series, we simply add the combined components:

$$\hat{y}_{t+h^{[1]}}^{[1]} = \bar{l}_{t+h^{[1]}} + \bar{b}_{t+h^{[1]}} + \bar{s}_{t-S+h^{[1]}}. \tag{7}$$

Note that if a component is missing in all K aggregation levels, then its contribution to $\hat{y}_{t+h^{[1]}}^{[1]}$ is zero. For instance, for a non-seasonal time series, the forecasted seasonal component is always zero for each k level, as well as for the combined $\bar{s}_{t-S+h^{[1]}}$.

4. Empirical evaluation

4.1. Experimental setup

In order to evaluate the proposed MAPA empirically, we compare its performance against ETS applied to the original data frequency directly. We use the annual, quarterly and monthly data from the M3-Competition (Makridakis & Hibon, 2000) and the semi-annual data from the Federal Reserve Bank of St. Louis, FRED.[2] The objective is to compare MAPA's performance with that of ETS at different time series frequencies. Table 2 shows the number of series available in each dataset, along with the number of observations in the holdout sample and the maximum aggregation level investigated in each frequency.

We have excluded the "Other" series of the M3-Competition, as their frequency is unknown. We retain a similar experimental setup to the M3-Competition. The forecast horizon for each dataset is equal to its holdout sample. This allows a direct comparison with the results and findings from the original competition (Makridakis & Hibon, 2000) and derivative works. For all datasets, with the exception of the annual, we set the maximum aggregation level so as to consider up to annual time series. For the annual dataset, we use one additional aggregation level.

We compare MAPA with state space ETS forecasts. ETS has been shown to perform very well on the M3-Competition dataset (Hyndman et al., 2002; Makridakis & Hibon, 2000). To fit the ETS, select the appropriate type and produce forecasts, we use the implementation available in the *forecast* package version 3.25 by Hyndman and Khandakar (2008) for the *R Project* (R Core Team, 2012). Note that the results obtained are slightly different from those of Hyndman et al. (2002), due to updates in the publicly available *R* code. We compare this with forecasts resulting from the mean and median combination of the forecasted components, as outlined in Section 3. At each level of aggregation, an ETS model is fitted. We produce two sets of MAPA forecasts, using the *Mean* and *Median* to combine the forecasted components at different aggregation levels. These will be referred to as *MAPA(Mean)* and *MAPA(Median)*. In order to assess the necessity of combining components instead of forecasts, we also benchmark MAPA against mean and median combinations of forecasts produced at each aggregation level. These will be referred to as *Comb(Mean)* and *Comb(Median)*.

Table 2
Data used for the empirical evaluation.

Original frequency	Source	Number of series	Holdout sample	Maximum aggregation level
Annual	M3-Competition	645	6	2
Semi-annual	FRED	1483	8	2
Quarterly	M3-Competition	756	8	4
Monthly	M3-Competition	1428	18	12

The performance of the proposed framework is measured using four different error metrics: (i) the Mean Percentage Error (MPE), which is appropriate for measuring the bias of the forecasts; (ii) the symmetric Mean Absolute Percentage Error (sMAPE), to allow comparisons with previous studies (Hyndman et al., 2002; Makridakis & Hibon, 2000); (iii) the Geometric Mean Relative Absolute Error (GMRAE); and (iv) the Mean Absolute Scaled Error (MASE), as proposed by Hyndman and Koehler (2006). These are calculated as:

$$\text{MPE} = m^{-1} \sum_{j=1}^{m} \frac{(y_j - \hat{y}_j)}{y_j}, \tag{8}$$

$$\text{sMAPE} = m^{-1} \sum_{j=1}^{m} \frac{2 |y_j - \hat{y}_j|}{(|y_j| + |\hat{y}_j|)}, \tag{9}$$

$$\text{GMRAE} = \sqrt[m]{\prod_{j=1}^{m} \frac{|y_j - \hat{y}_j|}{|y_j - \hat{y}'_j|}}, \tag{10}$$

$$\text{MASE} = m^{-1} \sum_{j=1}^{m} \frac{|y_j - \hat{y}_j|}{(n-1)^{-1} \sum_{r=2}^{n} |y_r - y_{r-1}|}, \tag{11}$$

where y_t and $\hat{y}_t$ are the actual and forecasted values, respectively, at time t; m is the number of observations in the hold-out sample; and n is the number of observations in the fitting sample. For GMRAE, a benchmark forecast $\hat{y}'_t$ is needed. For this purpose, we use conventional ETS forecasts, thus highlighting directly the way in which MAPA contributes to the forecasting accuracy of ETS.

4.2. Results

For each dataset, we provide results for the total forecast period, and short-term, medium-term and long-term forecasts. Table 3 presents the results for the annual M3-Competition dataset. Each column refers to a different error measure, and the best result for each error measure and forecast horizon is highlighted in boldface. Note that, since the maximum aggregation level for this dataset is two, the *Mean* and *Median* operators are the same for *MAPA* and *Comb*. Therefore, we summarise the results in a single row, under *MAPA* and *Comb* respectively. The results for *MAPA* and *Comb* are identical, as there is no seasonality in the annual time series.

Across the complete forecast horizon, the proposed MAPA outperforms the benchmark ETS for all error metrics apart from GMRAE. Considering sMAPE, it falls short of the best M3-Competition performance (Hyndman et al., 2002). Interestingly, if we look in more detail, the benchmark ETS clearly performs better for short forecast horizons, while

Table 3
M3 Competition results: annual time series.

	MPE %	sMAPE %	GMRAE	MASE
All ($t + 1$ to $t + 6$)				
ETS	−14.50	18.70	**1.00**	3.53
Comb	**−12.56**	**18.42**	1.03	**3.23**
MAPA	**−12.56**	**18.42**	1.03	**3.23**
Short ($t + 1$ to $t + 2$)				
ETS	**−4.30**	**10.60**	**1.00**	**1.51**
Comb	−4.83	11.17	1.15	1.57
MAPA	−4.83	11.17	1.15	1.57
Medium ($t + 3$ to $t + 4$)				
ETS	−14.30	19.50	1.00	3.46
Comb	**−12.82**	**19.03**	**0.97**	**3.21**
MAPA	**−12.82**	**19.03**	**0.97**	**3.21**
Long ($t + 5$ to $t + 6$)				
ETS	−25.10	25.90	1.00	5.62
Comb	**−20.04**	**25.07**	**0.97**	**4.92**
MAPA	**−20.04**	**25.07**	**0.97**	**4.92**

such is not the case for longer horizons. This provides some evidence that the proposed framework performs well for long-term forecasts.

Table 4 presents the results for the semi-annual dataset. Again, the maximum aggregation level is two, making the *Mean* and *Median* results identical, which are again summarised in single rows. A picture similar to that for the annual results emerges. Overall, ETS is outperformed across all error metrics. Both the bias and the forecasting accuracy are improved by the MAPA forecasts. The same is true for the Comb forecasts. The improvements are smaller for short forecast horizons, but become more substantial for longer horizons, providing additional evidence of the strong benefits for long-term forecasts. Although there are very small differences between the combination and MAPA forecasts, due to the small semi-annual seasonality, these cannot be seen in the summarised figures that are reported due to rounding.

Table 5 summarises the quarterly dataset. In this case, ETS performs best from $t + 1$ to $t + 8$, for all metrics but MASE. Considering the MASE, *MAPA(Mean)*, *MAPA(Median)* and ETS all perform the same. Looking at the results in more detail, one can see that both *MAPA(Mean)* and *MAPA(Median)* improve for longer horizons, to the extent that, for long-term forecasts, *MAPA(Mean)* ranks first according to sMAPE and MASE, while *MAPA(Median)* ranks first according to GMRAE. Both outperform ETS on all metrics apart from MPE, with ETS providing the least biased forecasts. Consistent with the results in Tables 3 and 4, MAPA performs better for long horizon forecasts than for short-term ones.

Comparing the forecast combination with MAPA, we can see that MAPA is consistently more accurate across all

Table 4
FRED database results: semi-annual time series.

	MPE %	sMAPE %	GMRAE	MASE
All ($t + 1$ to $t + 8$)				
ETS	−8.66	8.75	1.00	4.69
Comb	**−7.73**	**8.04**	**0.90**	**4.17**
MAPA	**−7.73**	**8.04**	**0.90**	**4.17**
Short ($t + 1$ to $t + 3$)				
ETS	−8.08	7.85	1.00	3.60
Comb	**−7.27**	**7.25**	**0.95**	**3.28**
MAPA	**−7.27**	**7.25**	**0.95**	**3.28**
Medium ($t + 4$ to $t + 6$)				
ETS	−9.33	9.20	1.00	5.08
Comb	**−8.24**	**8.38**	**0.88**	**4.45**
MAPA	**−8.24**	**8.38**	**0.88**	**4.45**
Long ($t + 7$ to $t + 8$)				
ETS	−8.50	9.40	1.00	5.72
Comb	**−7.66**	**8.73**	**0.88**	**5.07**
MAPA	**−7.66**	**8.73**	**0.88**	**5.07**

Table 5
M3-Competition results: quarterly time series.

	MPE %	sMAPE %	GMRAE	MASE
All ($t + 1$ to $t + 8$)				
ETS	**−2.86**	**9.84**	**1.00**	**2.40**
Comb(Mean)	−5.49	10.51	1.15	2.43
Comb(Median)	−5.14	10.68	1.16	2.44
MAPA(Mean)	−5.26	10.18	1.10	**2.40**
MAPA(Median)	−4.83	10.26	1.09	**2.40**
Short ($t + 1$ to $t + 3$)				
ETS	**−1.12**	**6.53**	**1.00**	**1.38**
Comb(Mean)	−3.60	8.10	1.34	1.67
Comb(Median)	−3.42	8.29	1.35	1.67
MAPA(Mean)	−3.21	7.62	1.26	1.64
MAPA(Median)	−2.97	7.68	1.24	1.62
Medium ($t + 4$ to $t + 6$)				
ETS	**−6.91**	**10.62**	**1.00**	2.58
Comb(Mean)	−9.57	11.16	1.10	2.58
Comb(Median)	−9.20	11.36	1.11	2.58
MAPA(Mean)	−9.37	10.86	1.05	2.55
MAPA(Median)	−8.89	10.95	1.05	**2.53**
Long ($t + 7$ to $t + 8$)				
ETS	**0.60**	13.62	1.00	3.64
Comb(Mean)	−2.20	13.15	0.97	3.35
Comb(Median)	−1.62	13.26	0.98	3.39
MAPA(Mean)	−2.17	**13.01**	0.95	**3.34**
MAPA(Median)	−1.51	13.08	**0.94**	3.36

Table 6
M3-Competition results: monthly time series.

	MPE %	sMAPE %	GMRAE	MASE
All ($t + 1$ to $t + 18$)				
ETS	−10.08	14.38	1.00	**2.12**
Comb(Mean)	−11.10	14.40	1.04	2.17
Comb(Median)	−10.40	14.90	1.06	2.17
MAPA(Mean)	−10.02	**13.89**	**0.99**	2.13
MAPA(Median)	**−9.01**	14.07	1.00	**2.12**
Short ($t + 1$ to $t + 6$)				
ETS	−7.51	12.01	**1.00**	**1.27**
Comb(Mean)	−7.57	12.42	1.13	1.43
Comb(Median)	−7.59	12.85	1.13	1.41
MAPA(Mean)	−6.49	**11.90**	1.07	1.38
MAPA(Median)	**−6.38**	12.01	1.07	1.36
Medium ($t + 7$ to $t + 12$)				
ETS	−7.17	13.69	1.00	**2.08**
Comb(Mean)	−5.91	13.73	1.01	2.12
Comb(Median)	−5.96	14.26	1.04	2.13
MAPA(Mean)	−5.89	**13.18**	**0.96**	**2.08**
MAPA(Median)	**−5.47**	13.34	0.99	**2.08**
Long ($t + 13$ to $t + 18$)				
ETS	−15.57	17.43	1.00	3.01
Comb(Mean)	−19.67	16.96	0.98	2.96
Comb(Median)	−18.06	17.50	1.01	2.98
MAPA(Mean)	−17.68	**16.59**	**0.95**	2.93
MAPA(Median)	**−15.20**	16.85	0.96	**2.92**

which achieved an sMAPE of 13.85% (Assimakopoulos & Nikolopoulos, 2000). For short-term forecasts, the relative performance of ETS improves, with only MPE and sMAPE favouring *MAPA(Mean)*. However, for medium- and long-term forecast horizons, both *MAPA(Mean)* and *MAPA(Median)* clearly outperform the benchmark ETS for all error metrics.

MAPA forecasts consistently outperform both *Comb (Mean)* and *Comb(Median)*. The seasonality in monthly time series is even more prominent, and the special handling of the seasonal component by MAPA is responsible for the improvements in accuracy. Moreover, the forecasts of *Comb(Mean)* and *Comb(Median)* are almost always worse than those of the benchmark ETS, demonstrating that a simple combination of different aggregation levels is not enough to achieve forecasting performance improvements.

Overall, across the four datasets, we can see that both *MAPA(Mean)* and *MAPA(Median)* improve upon ETS, particularly for long-term forecasts. Although the exact ranking of the competing methods depends on the error metric, it appears that the *MAPA(Mean)* is more accurate, while the *MAPA(Median)* is less biased. Across all datasets, both MAPA models are significantly better than the benchmark ETS for the medium, long and overall forecast horizons, as indicated by a Friedman test followed by an MCB post-hoc test, according to the procedure outlined by Koning, Franses, Hibon, and Stekler (2005). For short horizons, no significant differences were identified. In both the quarterly and monthly datasets that contain seasonal time series, MAPA significantly outperforms forecast combination across different aggregation levels. In many cases, the latter performs worse than the benchmark ETS.

horizons. Furthermore, *MAPA(Median)* is consistently less biased than *Comb(Median)*. The same is true for the *Mean* models. The improved MAPA results are attributed to the way in which seasonality, which is more prominent in the quarterly data, is handled by combining the components.

Table 6 provides the results for the monthly dataset. Considering the overall accuracy, *MAPA(Mean)* outperforms the benchmark ETS for all metrics but MASE. *MAPA(Median)* demonstrates a similar behaviour, achieving a similar performance to ETS for MASE. Note that, according to the sMAPE results, the performance of the proposed *MAPA(Mean)* is very close to that of the best method in the original M3-Competition, the Theta method,

5. Discussion

We find that MAPA improves on the forecasting performance (in terms of both accuracy and bias) of ETS. In particular, the good long-term performance of the MAPA can be explained by the proposed time series temporal aggregation. This step of MAPA functions as a low-pass filter that enhances the estimation of the low frequency components (level and trend). As a result, the long-term movements in the series associated with these low frequency components can be estimated better. In contrast, conventional ETS appears to perform best for short-term forecasts, where the impact of such low frequency components is small. The same improvements can be achieved by simply combining forecasts from different aggregation levels, as long as there is no seasonality.

In the second step of MAPA, as shown in Fig. 1, the time series of each level are smoothed using ETS. This allows a better estimation of the expected value of each time series component at each aggregation level. As was illustrated in the empirical results for the quarterly and monthly time series, a simple combination of the ETS forecasts is not adequate for seasonal time series. On the other hand, MAPA is able to provide performance improvements due to its selective consideration of the estimated components.

The third step combines the forecasted components, calculated at different temporal aggregation levels. Depending on the type of ETS model fitted at each aggregation level, the time series components may be either similar or different in their nature. The combination step can be seen as a further type of smoothing, a model smoothing, where the most common type of model across aggregation levels is weighted more heavily in the final forecast. This lessens the importance of model selection at each individual aggregation level. In contrast, the ETS forecasts are based on a single model that may be selected wrongly. The improved forecasting accuracy of MAPA is attributed to these multiple types of "smoothing".

MAPA requires an ETS model to be fitted at each aggregation level. This obviously increases the computations required. However, as the level of aggregation increases, the series derived have considerably fewer observations. Moreover, for many aggregation levels, no seasonal models are considered. These points limit the total computational time required. The calculations needed for the component transformations and their combinations are trivial, and add little to the required computations.

In Section 4, we focused on the improved accuracy of point forecasts; however, prediction intervals are also useful, in that they provide an indication of the uncertainty of the forecasts. To produce prediction intervals with MAPA, one can use a bootstrap method and simulate a large number of future sample paths, as described by Hyndman et al. (2002). This can be done using the final forecast of MAPA, as calculated in Eq. (7), thus making the process similar to calculating prediction intervals using simulation for conventional exponential smoothing.

In this paper, we discussed the calculation of predicted time series components using ETS. It is possible to consider alternative methods. Once forecasts for $\hat{y}_{i+h^{[k]}}^{[k]}$ have been constructed, we can use time series decomposition to extract each component from the forecasts (such as TRAMO SEATS, see Gómez & Maravall, 1996). In this case, the additive or multiplicative decomposition must be selected appropriately. Next, the resulting components must be transformed to additive.

6. A hybrid model

If we consider the way in which the forecasts for the *MAPA(Mean)* are constructed, we can see that $1/K$ of it is conventional ETS modelled on the original $Y^{[1]}$ time series. ETS performs better at short forecast horizons, because low frequency time series components do not have a prominent effect on the time series for such horizons. A combination of ETS with the *MAPA(Mean)* (or the *MAPA(Median)*) will potentially perform well for both short- and long-term forecasts. We construct simple combinations of ETS and MAPA forecasts using an unweighted average. The results of this combination are summarised in Table 7. The results for all four error metrics are provided, for the complete forecast trace, for each dataset. The numbers in brackets refer to the combination of ETS with the *MAPA(Median)*, while the rest refer to the combination of ETS with the *MAPA(Mean)*. Underlined figures outperform all ETS, *MAPA(Mean)* and *MAPA(Median)* values for that particular dataset and error metric from the respective Tables 3–6. In each case, the best result across all models and their combinations is highlighted in boldface.

For both the annual and semi-annual datasets, with maximum aggregation levels $K = 2$, the gains are small. In these cases, ETS does not contribute different information. On the other hand, we observe improvements for both the quarterly and monthly datasets. For all accuracy error metrics, sMAPE, GMRAE and MASE, the combined forecasting accuracy is superior to ETS, *MAPA(Mean)* or *MAPA(Median)* separately. The bias results for MPE demonstrate improvements over the benchmark ETS in most cases, but there are no cases in which the combined forecasts are the least biased. It is worth noting that, for the monthly dataset, the results for both the *MAPA(Mean)* and *MAPA(Median)* combinations with ETS outperform the Theta model that achieved the best performance in the original M3-Competition (Hyndman et al., 2002; Makridakis & Hibon, 2000).

Due to the transparency of MAPA, we can understand how these improvements in accuracy occur. Focusing on the *MAPA(Mean)* for the monthly dataset, we employed a maximum aggregation level of $K = 12$. Based on Eqs. (4) and (7), we can easily see, for example, that the ETS on the original time series is represented 13 times in the combined forecast for the level component, while aggregation levels 2 to K are represented only once. Essentially, the hybrid model is a weighted average of the components estimated at different temporal aggregation levels. We can generalise this by rewriting Eq. (4) as:

$$\bar{l}_t = \sum_{k=1}^{K} w_k l_t^{[k]}, \tag{12}$$

where $w_1 = (K + 1)/2K$ and w_2 to w_K are equal to $1/2K$. Eqs. (5) and (6) can be revised similarly. This finding re-

Table 7
Combined ETS with *MAPA(Mean)* and (*MAPA(Median)*).

	MPE %	sMAPE %	GMRAE	MASE
Annual	−13.56 (−13.56)	**18.37** (**18.37**)	1.01 (1.01)	3.34 (3.34)
Semi-annual	−8.20 (−8.20)	8.36 (8.36)	0.95 (0.95)	4.40 (4.40)
Quarterly	−4.06 (−3.84)	9.63 (**9.58**)	**0.99** (**0.99**)	2.27 (**2.25**)
Monthly	−10.05 (−9.55)	**13.69** (13.72)	0.96 (**0.95**)	**2.04** (**2.04**)

veals that temporal hierarchies could be considered which could improve the accuracy of the forecasts further.

7. Conclusions

We have proposed a novel forecasting framework that is based on using temporal aggregation to strengthen or attenuate different time series elements, thus allowing a more holistic modelling of time series. High frequency components are easier to model at low levels of aggregation, while low frequency components are more apparent at high levels of aggregation. This was the main motivation for this work.

A key innovation is the use of time series components instead of forecasts at each aggregation level. This has several computational advantages and improves the transparency of the proposed framework. More importantly, it allows us to determine how the different components are combined, which is useful for the seasonal component that may not appear at all aggregation levels. We showed that ETS is readily applicable to this task, and introduced transformations of the different ETS models to additive ones, simplifying the combination of the different components over the different aggregation levels.

The reconciliation of the forecasts was done by combining the forecasted components calculated at different temporal aggregation levels. We used the mean and median as combination operators, as they have been shown to be both robust and accurate. The proposed framework has been evaluated using four large datasets. We found a good forecasting performance relative to the benchmark ETS, which has been shown to perform well in a variety of forecasting applications, including, importantly, the M3-Competition datasets that we used in this case.

Fildes and Makridakis (1995) and Makridakis et al. (1982) stressed the need for more work on trend extrapolation and long-term forecasting. Although exponential smoothing has evolved substantially since then, long term forecasting has remained challenging. The proposed MAPA offers improvements in accuracy, in particular for long-term forecast horizons. We attribute this to the prominent impact of the low frequency time series components at long forecast horizons. Using MAPA, we improve the forecasting accuracy and bias of ETS, a widely researched and established method.

Trying to reconcile short- and long-term performances, we also demonstrated that weighted combinations are beneficial, hinting at temporal hierarchies that could improve forecasting performances further. In particular, for the monthly M3-Competition dataset, the forecasts derived from this process were more accurate than the best performer in the literature: the Theta model. We aim to explore such temporal hierarchies in future work, potentially leading to theoretically motivated combination weights.

Acknowledgments

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.ijforecast.2013.09.006.

References

Abraham, B. (1982). Temporal aggregation and time series. *International Statistical Review*, *50*, 285–291.

Agnew, C. E. (1985). Bayesian consensus forecasts of macroeconomic variables. *Journal of Forecasting*, *4*(4), 363–376.

Amemiya, T., & Wu, R. (1972). The effect of aggregation on prediction in the autoregressive model. *Journal of the American Statistical Association*, *67*, 628–632.

Andrawis, R. R., Atiya, A. F., & El-Shishiny, H. (2011). Combination of long term and short term forecasts, with application to tourism demand forecasting. *International Journal of Forecasting*, *27*(3), 870–886.

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., Song, H., & Wu, D. C. (2011). The tourism forecasting competition. *International Journal of Forecasting*, *27*(3), 822–844.

Babai, M. Z., Ali, M. M., & Nikolopoulos, K. (2012). Impact of temporal aggregation on stock control performance of intermittent demand estimators: empirical analysis. *Omega*, *40*(6), 713–721.

Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Operational Research Society*, *20*(4), 451–468.

Brewer, K. (1973). Some consequences of temporal aggregation and systematic sampling for ARMA and ARMAX models. *Journal of Econometrics*, *1*(2), 133–154.

Casals, J., Jerez, M., & Sotoca, S. (2009). Modelling and forecasting time series sampled at different frequencies. *Journal of Forecasting*, *28*, 316–342.

Cholette, P. A. (1982). Prior information and ARIMA forecasting. *Journal of Forecasting*, *1*, 375–384.

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

Clements, M. P., & Hendry, D. F. (2007). *An overview of economic forecasting* (pp. 1–18). Blackwell Publishing Ltd.

De Gooijer, J. G., & Hyndman, R. J. (2006). 25 years of time series forecasting. *International Journal of Forecasting*, *22*, 443–473.

Drost, F., & Nijman, T. (1993). Temporal aggregation of GARCH processes. *Econometrica*, *61*, 909–927.

Federal Reserve Bank of St. Louis (2012). Federal reserve economic data (FRED). http://research.stlouisfed.org/fred2/ (accessed: 20.11.2012).

Fildes, R., & Makridakis, S. (1995). The impact of empirical accuracy studies on time series analysis and forecasting. *International Statistical Review*, *63*(3), 289–308.

Gardner, E. S. (1985). Exponential smoothing: the state of the art. *Journal of Forecasting*, *4*(1), 1–28.

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

Gómez, V., & Maravall, A. (1996). *Programs TRAMO and SEATS: instructions for users*. (beta version: September 1996). Banco de España working papers, Banco de España.

Granger, C. (1993). Implications of seeing economic variables through an aggregation window. *Ricerche Economiche*, *47*, 269–279.

Hafner, C. (2008). Temporal aggregation of multivariate GARCH processes. *Journal of Econometrics*, *142*(1), 467–483.

He, C., & Xu, X. (2005). Combination of forecasts using self-organizing algorithms. *Journal of Forecasting*, *24*, 269–278.

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

Hughes, M. C. (2001). Forecasting practice: organisational issues. *Journal of the Operational Research Society*, *52*(2), 143–149.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, *27*(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., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach*. Berlin: Springer Verlag.

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.

Jose, V. R. R., & Winkler, R. L. (2008). Simple robust averages of forecasts: some empirical results. *International Journal of Forecasting*, *24*, 163–169.

Kolassa, S. (2011). Combining exponential smoothing forecasts using Akaike weights. *International Journal of Forecasting*, *27*(2), 238–251.

Koning, A. J., Franses, P. H., Hibon, M., & Stekler, H. O. (2005). The M3 competition: statistical tests of the results. *International Journal of Forecasting*, *21*(3), 397–409.

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

Makridakis, S., & Winkler, R. (1983). Average of forecasts: some empirical results. *Management Science*, *29*, 987–996.

Miller, C. M., Clemen, R. T., & Winkler, R. L. (1992). The effect of nonstationarity on combined forecasts. *International Journal of Forecasting*, *7*(4), 515–529.

Mohammadipour, M., & Boylan, J. E. (2012). Forecast horizon aggregation in integer autoregressive moving average (INARMA) models. *Omega*, *40*(6), 703–712.

Nikolopoulos, K., Syntetos, A. A., Boylan, J. E., Petropoulos, F., & Assimakopoulos, V. (2011). An aggregate–disaggregate intermittent demand approach (ADIDA) to forecasting: an empirical proposition and analysis. *Journal of the Operational Research Society*, *62*(3), 544–554.

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

R Core Team (2012). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/, ISBN 3-900051-07-0.

Silvestrini, A., & Veredas, D. (2008). Temporal aggregation of univariate and multivariate time series models: a survey. *Journal of Economic Surveys*, *22*(3), 458–497.

Souza, L. R., & Smith, J. (2004). Effects of temporal aggregation on estimates and forecasts of fractionally integrated processes: a Monte-Carlo study. *International Journal of Forecasting*, *20*(3), 487–502.

Spithourakis, G., Petropoulos, F., Babai, M. Z., Nikolopoulos, K., & Assimakopoulos, V. (2011). Improving the performance of popular supply chain forecasting techniques. *An International Journal of Supply Chain Forum*, *12*(4), 16–25.

Spithourakis, G., Petropoulos, F., Nikolopoulos, K., & Assimakopoulos, V. (2012). A systemic view of ADIDA framework. *IMA Journal of Management Mathematics* (forthcoming).

Syntetos, A. A., & Boylan, J. E. (2005). The accuracy of intermittent demand estimates. *International Journal of Forecasting*, *21*(2), 303–314.

Taylor, J. W. (2003). Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting*, *19*(4), 715–725.

Taylor, J. W. (2008). Exponentially weighted information criteria for selecting among forecasting models. *International Journal of Forecasting*, *24*, 513–524.

Timmermann, A. (2006). Forecast combinations. In G. Elliott, C. Granger, & A. Timmermann (Eds.), *Handbook of economic forecasting, vol. 1* (pp. 135–196). Elsevier.

Trabelsi, A., & Hillmer, S. (1989). A benchmarking approach to forecast combination. *Journal of Business and Economic Statistics*, *7*, 353–362.

Wei, W. (1990). *Time series analysis: univariate and multivariate methods*. Redwood City, CA: Addison-Wesley.

Weiss, A. A. (1984). Systematic sampling and temporal aggregation in time series models. *Journal of Econometrics*, *26*(3), 271–281.

Willemain, T. R., Smart, C. N., Shockor, J. H., & DeSautels, P. A. (1994). Forecasting intermittent demand in manufacturing: a comparative evaluation of Croston's method. *International Journal of Forecasting*, *10*(4), 529–538.

Nikolaos Kourentzes is with the department of Management Science at Lancaster University Management School (UK). He received his Ptychion degree from Athens University of Economics and Business (AUEB), and an M.Sc. in Management Science and a Ph.D. from Lancaster University Management School. His research focus is on time series prediction, with a particular emphasis on neural networks, input variable selection, high frequency data and intermittent demand.

Fotios Petropoulos is a Senior Research Associate in the Lancaster Centre for Forecasting of Lancaster University. He remained the Coordinator and Research Associate of the Forecasting & Strategy Unit of National Technical University of Athens for nine years, where he also completed his Doctorate in Engineering. Fotios' research interests include time series forecasting, optimal model selection, aggregation of data and policy forecasting. In addition, he is an expert in software analysis, system design and the development of integrated forecasting support systems.

Juan R. Trapero has been a lecturer at Universidad de Castilla-La Mancha since September 2010. He received the Ingeniero Industrial, M.B.A., and Ph.D. degrees from the Universidad de Castilla-La Mancha, Ciudad Real, Spain, in 2003, 2004, and 2008, respectively. He was working as a postdoctoral researcher from 2008 to 2010 in the School of Management, Lancaster University (UK), funded by a Marie-Curie Intra-European Fellowship. His research interests include algebraic identification and estimation in continuous-time of dynamical linear and non-linear systems and state space methods applied to time series, forecasting and control.