Cross-temporal aggregation: Improving the forecast accuracy of hierarchical electricity consumption

Spiliotis, Evangelos and Petropoulos, Fotios and Kourentzes, Nikolaos and Assimakopoulos, Vassilios

Forecasting and Strategy Unit, School of Electrical and Computer Engineering, National Technical University of Athens, 9 Iroon Polytechniou Str, 15773 Zografou Athens, Greece, School of Management, University of Bath, Bath, BA2 7AY, UK, Lancaster University Management School, Department of Management Science, Lancaster, Lancashire, LA1 4YX, UK

July 2018

manuscript No.
(will be inserted by the editor)

Cross-temporal aggregation: Improving the forecast accuracy of hierarchical electricity consumption

Evangelos Spiliotis · Fotios Petropoulos · Nikolaos Kourentzes · Vassilios Assimakopoulos

Received: 31-Jul-2018 / Accepted: -

Abstract Achieving high accuracy in load forecasting requires the selection of appropriate forecasting models, able to capture the special characteristics of energy consumption time series. When hierarchies of load from different sources are considered together, the complexity increases further; for example, when forecasting both at system and region level. Not only the model selection problem is expanded to multiple time series, but we also require aggregation consistency of the forecasts across levels. Although hierarchical forecast can address the aggregation consistency concerns, it does not resolve the model selection uncertainty. To address this we rely on Multiple Temporal Aggregation, which has been shown to mitigate the model selection problem for low frequency time series. We propose a modification for high frequency time series and combine conventional cross-sectional hierarchical forecasting with multiple temporal aggregation. The effect of incorporating temporal aggregation in hierarchical forecasting is empirically assessed using a real data set from five bank branches, demonstrating superior accuracy, aggregation consistency and reliable automatic forecasting.

Keywords Temporal aggregation · Hierarchical forecasting · Electricity load · Exponential smoothing · MAPA

Evangelos Spiliotis
Forecasting and Strategy Unit, School of Electrical and Computer Engineering, National Technical University of Athens, 9 Iroon Polytechniou Str, 15773 Zografou Athens, Greece
Tel.: +30-210-7723637
E-mail: spiliotis@fsu.gr

Fotios Petropoulos
School of Management, University of Bath, Bath, BA2 7AY, UK

Nikolaos Kourentzes
Lancaster University Management School, Department of Management Science, Lancaster, Lancashire, LA1 4YX, UK

Vassilios Assimakopoulos
Forecasting and Strategy Unit, School of Electrical and Computer Engineering, National Technical University of Athens, 9 Iroon Polytechniou Str, 15773 Zografou Athens, Greece

1 Introduction

Load forecasting encompasses a wide range of forecasting problems. Achieving high forecast accuracy can yield significant improvements to energy management and planning, leading to many economic and environmental benefits through energy conservation techniques such as load shifting, peak shaving and energy storing (Aslak Petersen et al., 2016; Barzin et al., 2015). Motivated by these potential gains, substantial work has been done to improve energy forecasting methods and models (Tratar and Strmčnik, 2016; Jurado et al., 2015; Trapero et al., 2015).

The literature focuses on three main classes of methods based on the prediction models used: the statistical, the engineering and the artificial intelligence methods. A review on the load forecasting methods in the building sector has been conducted by Zhao and Magoulès (2012) and Suganthi and Samuel (2012), who looked at load modeling for the electricity production sector. They find that each forecasting method has its own strengths and weaknesses with respect to the problem at hand, available data and the level of acceptable complexity. The literature does not identify a best method and therefore the model selection problem remains an unresolved key modeling issue. This is particularly relevant for practice, where reliable selection of forecasts is desirable.

To mitigate this problem, in the general time series forecasting context, Kourentzes et al. (2014) proposed using multiple temporal aggregation (MTA), later generalized by Athanasopoulos et al. (2017). This approach is based on temporally aggregating a time series at multiple levels, which transforms the original data to lower time frequencies, highlighting different aspects of the series (Pedregal and Trapero, 2010). Mainstream time series modeling literature has mainly focused on identifying the single optimal level that makes modeling simpler (Silvestrini and Veredas, 2008). On the other hand, MTA models the series at multiple aggregation levels and combines the resulting forecasts. This has two key advantages: it provides a holistic modeling approach, focusing at both high and low frequency components that are highlighted at different temporal aggregation levels; and it mitigates modeling uncertainty, since the final forecast is not based on a single forecasting model.

Mitigating modeling uncertainty is crucial when dealing with hierarchies and forecasting several connected time series. In most problems related with energy conservation, management and pricing, any decision taken is multi-layered, considering and affecting multiple levels of the energy system it refers to. For example, when optimizing energy use of a building (top level), the individual energy uses (heating, cooling, lighting etc.) must also be taken into consideration (bottom level). These decisions are usually supported by forecasting systems, which produce forecasts for all levels of the hierarchy. From the one hand, this requires model selection and estimation for several time series and from the other hand it has the undesirable consequence that lower level forecasts may not sum up to the higher level forecasts and vice-versa, as they are produced by independent forecasting models. Forecasts in this case need to be reconciled to ensure aggregation consistency across levels, as otherwise decisions taken at different levels will not be aligned. To this end, cross-sectional hierarchical forecasting methods, such as "bottom-up" and "top-down", have been proposed to achieve reconciliation.

Given the need for applying cross-sectional approaches to the problem at hand, the question arises whether Multiple Temporal Aggregation (MTA) could be exploited to mitigate modeling uncertainty and improve forecasting performance, in a similar fashion to what has been reported in the literature for low frequency time series forecasting problems. This realized using the Multiple Aggregation Prediction Algorithm (MAPA) by Kourentzes et al. (2014). We propose a modification to make MAPA appropriate for high frequency time series and an approach to combine conventional cross-sectional hierarchical forecasting with MAPA. Our results show that our approach contributes towards *(i)* decreasing model uncertainty and increasing accuracy while *(ii)* ensuring reconciled forecasts across the hierarchy. Both enable automation of forecasting in such problems, aiding decision makers.

The rest of the paper is organized as follows: in the next section we review the work done so far in the temporal and cross-sectional aggregation literature. Section 3 describes our methodological approach including the methods used. The data of our case study and the experimental set-up is in section 4. Section 5 presents the results, followed by concluding remarks in section 6.

2 Literature Review

2.1 Cross-sectional hierarchies for forecasting

Energy applications are closely related with hierarchical structures and their accurate extrapolation. From supervision and management to pricing, energy conservation and storing, managers must consider diverse information from various levels of their systems to make the right decisions and act proactively. Advances in data collection, using innovations such as smart meters, further promote the need of exploiting the information hidden in hierarchies.

In order to be meaningful and consistent, the forecasts at higher levels must be equal to the sum of the individual lower level forecasts that make up the respective higher levels. The literature has investigated a variety of cross-sectional hierarchical approaches that are able to produce reconciled forecast (Athanasopoulos et al., 2009). The “bottom-up” approach aggregates forecasts of the lowest level of the hierarchy to obtain forecasts of all higher levels, while the “top-down” approach disaggregates the top-level forecasts to obtain the forecasts for the lower levels (Gross and Sohl, 1990). Another alternative is the middle-out approach, where the forecasts are produced at a middle level and are then aggregated or disaggregated as needed. Recently, Hyndman et al. (2011) introduced the “optimal combination” approach where all series of the hierarchy are forecasted independently and are subsequently combined using a regression model.

There is no consensus in the literature as to which approach is superior. The top-down approach is considered to be more appropriate for long term forecasts (Shlifer and Wolff, 1979), as it effectively captures the trend of the data (D’Attilio, 1989). On the other hand, the bottom-up approach performs better among highly correlated time series (Dangerfield and Morris, 1992) as it it highlights the special characteristics of the disaggregated data (Gordon et al., 2000; Shlifer and Wolff, 1979), while it also leads to less biased and more robust forecasts, at least when reliable and non-missing data are present at the lowest levels (Schwarzkopf et al., 1988). The correlation of the individual time series and their errors (Zotteri et al., 2005), as well as their variability (Tiao and Guttman, 1980; Kohn,

1982) might indicate which approach is preferable. However, the performance differences between the two approaches can often be minimal in practice, displaying also insignificant advantages over other formal or informal strategies reported in the literature (Fliedner and Lawrence, 1995). For instance, Widiarta et al. (2008) showed that their performance was nearly identical when forecasting item-level demands that followed a first-order moving average process or a first-order univariate autoregressive one, at leasts for low lag1 autocorrelation values (Handik et al., 2007). Information from all hierarchical levels could be considered instead, with evidence of benefits for the overall forecasting performance (Athanasopoulos et al., 2009; Hyndman et al., 2011).

In the field of load forecasting both top-down and bottom-up approaches are used for energy planning and management (Chalal et al., 2016; Kavgic et al., 2010). However, the latter is more popular (Heiple and Sailor, 2008), given that energy models usually correlate load with temperature data, which are monitored at the lowest levels of the hierarchy. Recently, Lai and Hong (2013) investigated the performance of various approaches for improving forecasting accuracy in electric usage by considering a geographic hierarchy. They showed that: (i) at lower levels the average of temperatures from multiple weather stations provides the best representation of weather, (ii) at upper levels the data sample strongly influences the modeling preferences and (iii) top-down and bottom-up approach display similar performance at the top level of the hierarchy.

2.2 Multiple temporal aggregation for forecasting

Energy forecasting deals with modeling challenges related to various interconnected uncertainties: sampling, parameter and model. Limited sample may obscure the underlying structure of the observed series and affect parameter estimation. This in turn can change the identified model structure, even if we assume that the appropriate model family is chosen, which itself is uncertain (Zhao and Magoulès, 2012). Instead, MTA can be used to mitigate the need to identify a single 'correct' model, or rely on a unique estimation of parameters.

MTA is based on temporal aggregation of time series. Silvestrini and Veredas (2008) studied the effect of temporal aggregation in the forecasting performance of univariate and multivariate time series models and provided evidence of performance improvement. They found that although there are merits in using temporal aggregation, but it is difficult to identify the optimal temporal aggregation level. Weiss (1984) offered insights of its impact in econometric models by considering the relationships between variables, reaching similar findings. In brief, temporal aggregation simplifies the identifiable structure and lessens the noise component of the series, yet depending on the aggregation level it may be that too much information has been filtered and therefore the resulting forecasts are of inferior quality. In a supply chain context, for slow moving items, temporal aggregation works as a "self-improving mechanism" (Nikolopoulos et al., 2011; Spithourakis et al., 2011) by revealing patterns which are more clear in lower frequencies. Yet, the difficulty in identifying the optimal aggregation level and selecting an appropriate model remains an issue (Petropoulos and Kourentzes, 2015).

In this respect, MTA, which instead of choosing a single level is aggregating series to multiple lower frequencies and combining the individual forecasts produced per level, becomes very promising (Petropoulos and Kourentzes, 2014). Given that at lower aggregation levels periodic components, such as seasonality, are dominant and that at higher levels these are filtered to reveal long-term ones, such as trends, every single level has valuable information to offer (Kourentzes et al., 2014). This is particularly relevant to fast-moving data, such as load forecasting applications, where the high sampling frequency displays increased noise and introduces multiple seasonal patterns at the original sampling frequency, which require data preprocessing of high complexity (Dudek, 2016).

MTA was proposed by Kourentzes et al. (2014) as implemented in MAPA, although the term itself was coined by Petropoulos and Kourentzes (2014) as a more general concept. Since, different variations of MTA have appeared, notably the Temporal Hierarchies (Athanasopoulos et al., 2017) and specialized variants for intermittent demand (Petropoulos and Kourentzes, 2015) and promotional modeling (Kourentzes and Petropoulos, 2016). MAPA models multiple aggregated views of a time series, using independent exponential smoothing models. The resulting outputs of the models from each individual aggregation level are then combined to produce a final forecast. The key advantages of the approach is that by using a different model per frequency, different time series components are captured, as these are differentially highlighted in different temporal aggregation levels; and modeling uncertainty is mitigated leading to performance gains, due to the multiple modeling views. The improvements have been reported both for short and long term forecast horizons, across different applications (Kourentzes and Petropoulos, 2016; Barrow and Kourentzes, 2018). Kourentzes et al. (2017) showed that although MAPA is not optimal at any aggregation level, it still provides more accurate forecasts than conventional approaches to temporal aggregation, as it is very resistant to any modeling misspecification.

Although a lot of research has been undertaken in the direction of accurately forecasting and reconciling energy related hierarchical time series, limited work has been done to address the increased modeling uncertainty that arises. In this respect we investigate how to effectively deal with model uncertainty in complex energy consumption hierarchies of high frequency data, while maintaining good forecasting performance. The proposed methodological approach combines these so far separate aggregation frameworks, cross-sectional and temporal, to gain reconciled forecasts of reduced modeling complexity.

3 Methodology

In this section we describe the proposed methodology to merge cross-sectional hierarchies and MAPA. We first describe the individual methods and then proceed to describe the encompassing methodology.

3.1 Aggregation and forecasting methods

Cross-sectional and temporal aggregation will be combined within the framework so as to achieve both reconciled forecasts and reduced modeling uncertainty. The aim is to provide a solution that will be reliable and automatic in a practical setting.

3.1.1 Hierarchical forecasting

Regardless of the forecasting methods used to extrapolate the electricity consumption time series for the different levels of the hierarchy, the individual forecasts must be reconciled to be useful for any subsequent decision making.

First we introduce the necessary notation. Let k denote the level of the hierarchy. Level 0 refers to the completely aggregated series, while level K the most disaggregated time series. m_i denotes the total number of series at level i, $i = 0, 1, 2, \ldots, K$ and $m = m_0 + m_1 + \ldots + m_K$ denotes the total number of series in the hierarchy. Let $Y_{x,t}$ denote the value of the x^{th} series at time t. Y_t represents the aggregate of all series at time t, $Y_{i,t}$ the value of the i^{th} series of level 1 at time t, $Y_{ij,t}$ the value of the j^{th} series used to disaggregate series $Y_{i,t}$ at time t; and so on. Vector $\mathbf{Y_{i,t}}$ denotes all observations at level i and time t such as $\mathbf{Y}_t = [Y_t, \mathbf{Y}_{1,t}, ..., \mathbf{Y}_{K,t}]$. Similarly, $\hat{Y}_{x,n}$(h) denotes the h-step-ahead forecasts of series x, also known as base forecasts. $\hat{\mathbf{Y}}_n(h)$ denotes the vector consisting of the base forecasts and $\tilde{\mathbf{Y}}_n(h)$ the vector consisting of the final hierarchical forecasts. Finally, $\mathbf{S}$ is a 'summing' matrix of order $m \times m_K$ used to aggregate the lowest level series so that $\mathbf{Y}_t = \mathbf{S}\mathbf{Y}_{K,t}$. The top row of $\mathbf{S}$ is a unit vector of length m_K, the bottom section is an $m_K \times m_K$ identity matrix, while the middle parts are vector diagonal rectangular. Matrix $\mathbf{S}$ gives a numeric representation of the hierarchical structure.

Cross-sectional hierarchical reconciliation can be expressed as a linear combinations of the unreconciled base forecasts. Using the notation above: $\tilde{\mathbf{Y}}_n(h) = \mathbf{SP}\hat{\mathbf{Y}}_n(h)$, where $\mathbf{P}$ is an appropriate matrix of order $m_K \times m$ and $\tilde{\mathbf{Y}}_\mathbf{n}(\mathbf{h})$ are the reconciled forecasts. All approaches that are widely used in the literature, bottom-up, top-down and optimal combination, can be expressed in these terms, differing only on the specification of $\mathbf{P}$.

The *bottom-up* approach aggregates the forecasts of the lowest level of the hierarchy $\hat{\mathbf{Y}}_K(h)$ to obtain the forecasts of the higher levels. This is done by simply summing the base forecasts from the lowest to the highest levels of the hierarchy according to its structure. In this respect, for the bottom-up approach $\mathbf{P} = [\mathbf{0}_{m_K \times (m-m_K)} \mid \mathbf{I}_{m_K}]$, where $\mathbf{0}_{l \times k}$ is a null matrix of order $l \times k$ and I_k is an identity matrix of order $k \times k$.

Next, the *top-down* approach disaggregates the forecasts of the highest hierarchical level $\hat{\mathbf{Y}}_n(h)$ to obtain the forecasts of the lower levels based on historical proportions of the data. For this approach $\mathbf{P} = [\mathbf{p} \mid \mathbf{0}_{m_K \times (m-1)}]$, where $\mathbf{p} = [p_1, p_2, ..., p_{m_K}]$ is a vector of proportions that sum to one. In the present study we use the average historical proportions p_j for implementing the method.

$$p_j = \sum_{i=1}^{n} \frac{y_{j,t}}{Y_t},$$

where p_j reflects the average of the historical proportions of the j=1,...,k bottom level series $Y_{j,t}$ over the period t=1,,n relative to the total aggregate Y_t. Other alternatives are the use of the proportions of the historical averages and the forecasted proportions, as described by Athanasopoulos et al. (2009).

Finally, the *optimal combination* identifies $\mathbf{P}$ so as to provide the minimal reconciliation errors, i.e. enforce aggregation consistency across forecasts by requiring only minimal changes of the base forecasts. Hyndman et al. (2011) shows that in this case $\mathbf{P} = (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'$, which implies that it depends only on the structure of the hierarchy. Note that this formulation also implies that forecasts from all time series are linearly combined, in contrast to only the lower or top levels, as prescribed by the bottom-up and top-down. Therefore, more information is retained by the optimal combination reconciliation, but also requires reasonable forecasts for all time series of the hierarchy.

3.1.2 Multiple Aggregation Prediction Algorithm

The MAPA algorithm can be separated into three steps: aggregation, forecasting and combination. Starting with temporal aggregation, let Y be a time series of periodicity f and length n and y_t denote its observation at point t. We can temporally aggregate Y by summing the values of the series at the original frequency y_t in buckets of length k. The temporally aggregated time series created $Y^{[k]}$ has n/k observations with values

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

For example, given a monthly time series with periodicity $f = 12$, we get the original series for $k = 1$, a quarterly series for $k = 3$, a half-annual series for $k = 6$ and an annual series for $k = 12$. We can apply temporal aggregation for any value of $k \leq n$, although in practice we do so for $k \ll n$ in order for $Y^{[k]}$ to have enough observations for fitting a forecasting model. We also note that if the remainder of the division n/k is not zero, we remove $n - [n/k]$ observations from the beginning of the series in order to form complete aggregation buckets.

Following temporal aggregation, a prediction model is fit to each of the created series. Although in its original form MAPA was proposed using the complete family of exponential smoothing, the selection of the forecasting model is up to the practitioner and may depend on the type of the data, the application and available resources. The substantive issue here is that instead of handling each forecast as a single value, we decompose it into three basic components: level (l_i), trend (b_i) and seasonality (s_i). This is done to combine the individual components instead of forecasts, which is useful as at each temporal aggregation level a different model can be fit and combining by components allows to draw only the necessary information from each level.

In its third step, MAPA combines the components estimated per aggregation level to produce the final forecast. This can be done using a variety of combination operators, such as the mean or median. In this work we consider the median since it is less affected by poorly estimated components due to extreme values and other types of outliers, noise and limited training sample, and can therefore lead to more robust forecasts. This can become extremely helpful when dealing with noisy data of high frequency (like hourly energy consumption time series), where even outlier detection methods are possible to fail or under-perform. The final h-step-ahead forecast of the series is calculated as:

$$\hat{y}_h = \bar{l}_h + \bar{b}_h + \bar{s}_h, \tag{2}$$

where $\bar{l}_h$, $\bar{b}_h$ and $\bar{s}_h$ is the median of the level, trend and seasonal components estimated for the h-step-ahead forecast produced across all aggregation levels considered. We note that in order to combine the forecasts, all the components must first be transformed into an additive form for (2) to hold, irrespective of the type of model used. Multiplicative components can be transformed to additive easily by multiplying them with the respective level. Additionally, if a component has not been estimated for an aggregation level (e.g. in case of non-seasonal time series or use of non-trended models), we set it equal to zero. The reasoning behind this is simple: as MAPA does not assume knowledge of the true process, if at a level a trend is identified, but at another none is identified and set to zero, we have no indication to prefer one or the other option. Therefore these are combined into a damped trend. Naturally, if most levels identify zero trend, then any estimated trend will be diminished and vice-versa.

In forecasting load data there is a crucial consideration that should not be overlooked: accurate prediction of peaks is important. Peak consumption is strongly correlated to variables such as energy prices and system stability. However, when applying temporal aggregation on time series, the produced forecasts will be much smoother than the original data due to (1) that acts as a moving average filter. Also, any subsequent combinations across temporal aggregation levels will exhibit damped seasonality.

Hourly energy data typically exhibit strong daily and weekly seasonality. There is a load profile that occurs every 24 hours, capturing the day-night cycle, and every 168 hours, capturing the different days of the week cycle, and particularly the difference between work-days and weekends. These long seasonal periodicities permit to consider multiple temporal aggregation levels that can potentially exhibit seasonality, specifically: 2, 3, 4, 6, 7, 8, 12, 14, 21, 24, 28, 42, 56 and 84. As a result, the peak load will be poorly forecasted, due to the shrinkage of the seasonal component imposed by temporal aggregation.

A solution that keeps seasonality unaffected is to apply temporal aggregation on the seasonally adjusted data and re-seasonalise the final forecasts. That way a deterministic seasonality is forced helping us to effectively handle the peaks. An example of this phenomenon and the proposed solution is provided in figure 1, where the hourly energy demand of a commercial building is forecasted for five days ahead. As shown, MAPA produces forecasts with shrunk seasonal indexes, while MAPA on a seasonally adjusted series maintains the original seasonal pattern of the data.

Obviously this approach makes using the full exponential smoothing family unnecessary, as seasonality is modeled externally (in (2), $\bar{s}_h = 0$). We impose a further simplification: in the decision relevant forecast horizons (1 to 7 days ahead) consumption data do not exhibit persistent trends, as the effect of possible behavioral changes or operational adjustments is impossible to be captured within such short periods. Therefore, we only consider the level variant of exponential smoothing (in (2), $\bar{b}_h = 0$), which is the widely used Simple Exponential Smoothing (SES) (Gardner, 1985). In this regards the final forecast of MAPA will be the median of the levels calculated.

Undoubtedly, if the same forecasting method is to be applied to all temporally aggregated views of a series, there is no reduction of the model selection uncertainty. However, MAPA still provides benefits in terms of mitigating the parameter estimation uncertainty, as the method parameters are estimated on multiple views

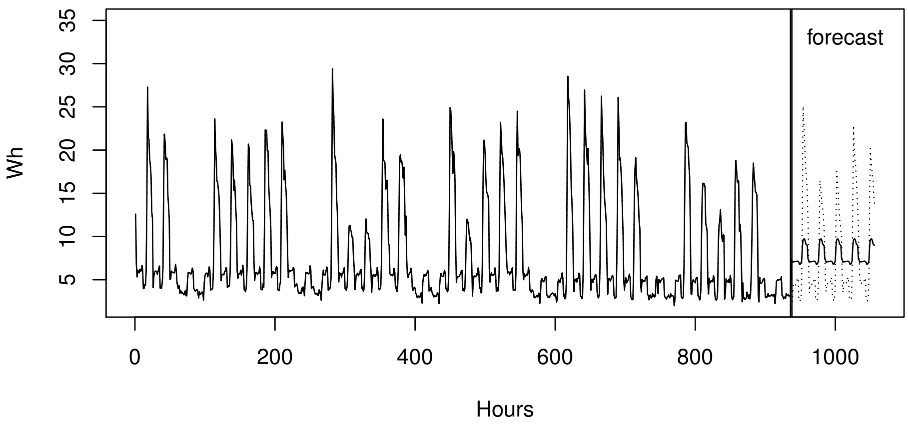

Fig. 1 The effect of MAPA (continuous) on an hourly electricity consumption of a commercial building with strong seasonality. In contrast to seasonally adjusted SES (dotted), seasonal indexes produced are significantly shrunk.

of the series. SES produce forecasts using a single estimation of the smoothing parameter and initial level state. Both of those parameters are specified through appropriate criteria. However there is always the risk of poor parameterization due to the effect of outliers and other unusual values, especially for series of high frequency, where noise may still be dominant. By calculating these parameters multiple times across temporally aggregated series, we can significantly reduce the modeling uncertainty and increase the robustness of the model.

An alternative solution to the seasonality shrinkage of MAPA can be achieved by using a weighted combination. The final components $\bar{l}_h$, $\bar{b}_h$ and $\bar{s}_h$ in (2) are the result of the unweighted combination of the components estimated at each aggregation level. Although for both level and trend the long-term dynamics, as captured by the higher levels of temporal aggregation, enrich them, for the seasonal component, $\bar{s}_h$, it can lead to undesired shrinkage. We propose to mitigate this shrinkage using a simple weighting scheme: each aggregation level k is weighted by $1/k$, effectively lessening the shrinkage. The combination for both level and trend components remains unweighted. Kourentzes et al. (2014) identified this shrinkage effect and proposed a weighed combination for relatively low frequency (up to monthly) time series, to mitigate this. The weighting scheme we propose is more aggressive in retaining the high frequency aspects of the seasonal pattern, which are crucial for high frequency time series forecasting. Note that eliminating shrinkage altogether is not desirable, as it has been shown to be beneficial (Miller and Williams, 2003).

The decomposition approach simplifies the specification of MAPA substantially, considering both the number of parameters to be estimated (in exponential smoothing the biggest estimation cost comes from the seasonal component) and the number of possible alternative exponential smoothing models considered at each aggregation level. Both will result in substantial speed-ups in model specification, and potential accuracy gains, particularly when the in-sample data are limited in length. On the other hand, the weighted combination approach avoids imposing a specific decomposition, which may be erroneous, and does not require sequential estimation, of the decomposed seasonal profile and then the MAPA fit, that can introduce modeling bias. Finally, it does not restrict MAPA to a single exponential smoothing model type, hence mitigating both estimation (like the decomposition alternative) and model selection uncertainty. In any case, both of the modifications proposed for MAPA to better deal with high-frequency data display multiple advantages over its originally proposed form, leading potentially to improvements in forecasting performance.

3.1.3 Exponential smoothing

The Single Exponential Smoothing (SES) model is used to produce the benchmark forecasts, when no temporal aggregation is used. It is also used to produce the individual forecasts for each temporally aggregated view of the time series. The model is used to track the local level of a given series by inspecting its changes over time and is expressed through the following equations:

$$\begin{aligned} \hat{y}_{t+1} &= l_t, \\ l_t &= l_{t-1} + \alpha e_t, \\ e_t &= y_t - \hat{y}_t, \end{aligned} \tag{3}$$

where l_t is the estimated level of the series and $\hat{y}_t$ the forecast of SES at point t. α is the smoothing parameter used for adjusting the running level of the series and can take any value between 0 and 1. In case $\alpha = 1$, SES becomes equal to the naive method, while if $\alpha = 0$ the produced forecasts are equal to l_0, the value of the initial level. In general, the higher the value of α, the more weight is assigned to the more recent observations in calculating the level.

In order to estimate the model we first specify the values of l_0 and α. This is done by maximising the likelihood $\mathcal{L}$ of the model (Hyndman et al., 2002):

$$\mathcal{L}(\alpha, S_0) = -\frac{n}{2} log(\sum_{t=1}^{n} (e_t)^2),$$

where n is the length of the series and the error e_t is conditional on the smoothing parameter α and the initial state l_0 used. This criterion is utilized within the study to individually optimize the parameters of the model across all the series of the hierarchy.

A seasonal variant of the model can be easily constructed by including a seasonal component. The same stands for the case of trend. All typical variants of exponential smoothing are described by Hyndman et al. (2002). In this paper we focus only on the additive approaches, that may allow for trend and seasonality. Note that the additive formulation of exponential smoothing is more robust to time series with very low or zero values, which can be the case for the disaggregated building electricity consumption time series.

3.2 Forecasting methodology

When dealing with real data it is common that there may be issues, such as data collection errors. The reasons for obtaining abnormal data vary and can be metering and data streaming problems, outages, failures of the electricity provider's system, and so on. These can have an adverse impact on the performance of the forecasting system, due to the carry-over effect of the outliers on the forecasts and the bias introduced in the estimates of the model parameters (Ledolter, 1989). Therefore, data cleansing becomes a task of significant importance.

Missing values are imputed to enable further analysis and modeling. Given a missing value X_t at point-hour t, the arithmetic mean of the observations X_{t+168} and X_{t-168} is used as its replacement to take into account both the weekly and hourly seasonality of energy consumption (since X is an hourly series of both daily and weekly cycles, seasonal effects are theoretically repeated every 7days*24hours=168hours). If observation X_{t+168} is unavailable, X_{t-168} is used as a replacement while, for the rest of the cases, a simple linear interpolation between the last respective known and the next available observations is applied to estimate the missing values. The imputed observations are used both for model estimation and evaluation so that more representative results are obtained.

Another important data consideration is special days, such as bank holidays, which can affect the forecasting performance negatively (Erişen et al., 2017). These can reduce accuracy during both outlying and normal periods. Barrow and Kourentzes (2018) evaluated various approaches to deal with these and found that for conventional forecasting methods, such as SES, one of the best performing approaches is to correct them. Therefore, we consider additive outliers and level shifts using the detection approach proposed by Chen and Liu (1993). Additive outliers adjustments will be used to mitigate the effect of extreme values, while level shift adjustments will deal with temporal changes on the level of the series due to outages, change in equipment and technical problems.

The individual time series of the hierarchy are then seasonally adjusted to effectively capture the consumption peaks. As discussed in section 3.1.2. Deseasonalization is performed by means of classical decomposition by moving averages (Kendall and Stuart, 1983), with a seasonal periodicity of 168 hours. We use additive decomposition, so as to avoid any complications with very low demand values at the most disaggregated level:

$$Y_t = b_t + s_t + e_t,$$

where b_t, s_t, and e_t denote the component of trend, seasonality and error, respectively. In order to estimate b_t, a moving average of order equal to the periodicity of the data is applied and then used to remove the trend from the original series. The seasonal component is computed by averaging for each time unit over all periods, then centering. Finally, the error component is the remainder of the original time series when b_t and s_t are removed.

Alternative seasonal cycles, such as 24, were also tested, but rejected due to the impact of working and non-working days, resulting in less homogeneous seasonal profiles, as evident by the corresponding seasonal plots, see figure 2. The classical additive decomposition is applied to the time series for alternative periodicities (24 and 168 hours) and the extracted seasonal component is plotted against the individual periods in the season. In this respect, periods of low variance indicate strong seasonal patterns, and vice versa. Observe that the weekly pattern has substantially lower variance than the daily one, indicating that the former is estimated more accurately and is preferable to the daily one.

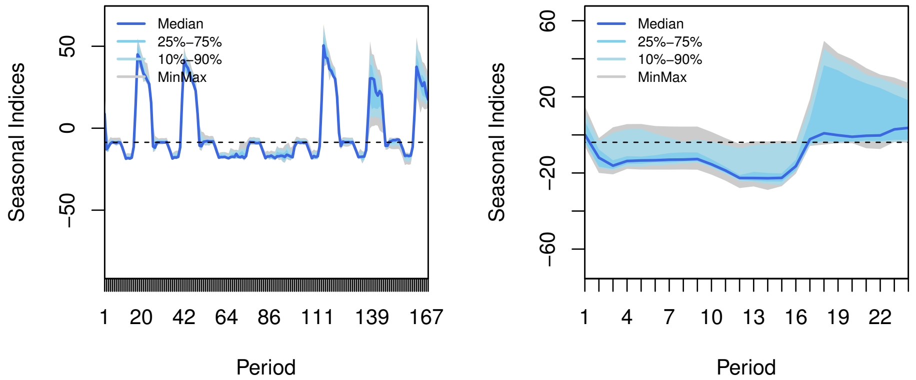

Fig. 2 Distribution of seasonal indices for the total electrical consumption of the bank branches for seasonal cycles of 168 (left) and 24 (right) hours. Given that in a time series with strong seasonality the observations will be overlapping, we anticipate to low variance around the seasonal profile. This is evident for the weekly profile, while for the daily profile differences between working days, weekends and bank holidays introduce substantial variance.

Data transformations, such as the Box-Cox one, which could have been used to normalize the raw data, simplify their patterns and enhance forecasting performance (Beaumont, 2014) were not considered in the present study. This is because many of the time series examined display values close to zero, making their implementation ineffective. Transformations are not applicable either after seasonally adjusting the data since additive decomposition may lead to time series of negative values. Once the data pre-processing is complete, each time series is forecasted using MAPA. The resulting forecasts are re-seasonalised, using the seasonal indices estimated before. After producing the forecasts, these are reconciled across the various levels of the hierarchy. As the literature is inconclusive as to best cross-sectional aggregation approach, we retain all and evaluate the best one. An overview of the proposed methodology is presented in figure 3.

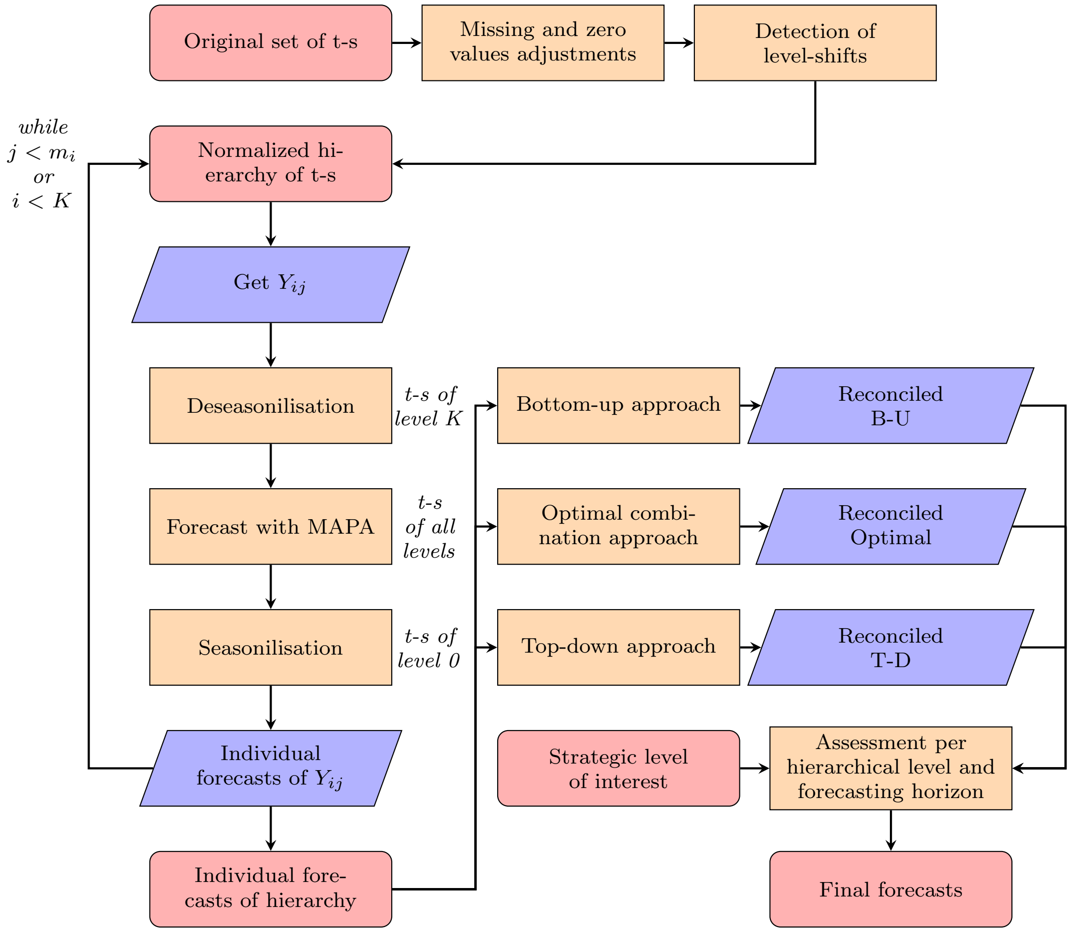

Fig. 3 Flow chart of the proposed methodological framework applied to a K-level hierarchy.Here, Y_{ij} indicates the j_{th} time series of level i.

4 Experimental Design

4.1 Data and case study

The proposed methodology is applied on a group of five bank branches located in Athens, Greece. We examine the benefits in terms of accuracy, complexity and decision support.

The bank branches form a three-level hierarchy representing per level the total energy needs of the bank (level 0), the energy consumption per bank branch (level 1) and end use (level 2): Heating, Ventilation and Air Conditioning (HVAC), devices connected to UPSs (cameras and safes) and Lighting. The structure of the hierarchy is presented in figure 4, while a typical example of the time series of each level is provided in figure 5. The available data (energy consumption in kWh) span for 9.5 weeks (1612 hourly observations, from 5-Jan-12 to 12-March-12). Missing observations account for about 2% of the whole sample, while special days for about 3%. The majority of them belong in the train sample.

Note that the relatively small size of the dataset is another challenge that needs to be tackled among the others discussed, i.e., the high-dimensional seasonal profile, model and parameter uncertainty, missing values and special days. Given that the methods typically used in such applications, such as neural networks, strongly rely on extended samples of data, generating robust forecasts through alternative approaches like the one proposed becomes vital. For instance, it would be interesting to see whether our approach effectively captures seasonality, ensures reliable parameter estimations and leads to accurate forecasts, even when relatively long horizons are considered. If that is the case, then this would be an additional strength of the proposed framework.

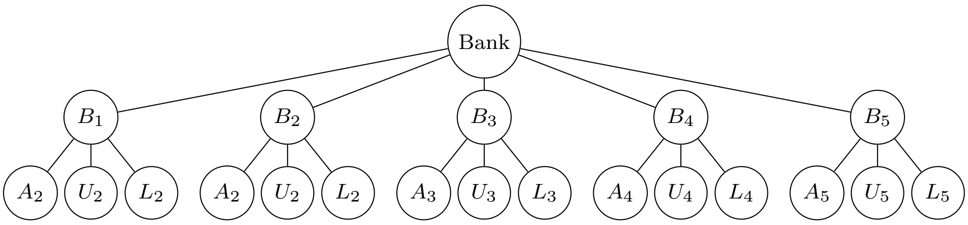

Fig. 4 The three-level hierarchical tree diagram of the bank case-study. Here, B_i, A_i, U_i and L_i stand for the i_{th} **B**ranch, HV**A**C, **U**PS and **L**ighting energy use, respectively.

4.2 Experimental setup

The forecasting performance of the methods will be measured by producing forecasts at all the levels of the hierarchy and across different horizons to indicate per level possible gains for short- (1-48 hours ahead, up to 2 days), medium- (49-120 hours ahead, up to 5 days) and long-term (121-168 hours ahead, up to a week) decisions. Thus, the most appropriate combination of temporal and cross-sectional aggregation methods will be empirically demonstrated. The forecast horizons mirror the current bank's energy manager practices.

At the beginning of every week forecasts are produced for all branches to highlight possible threats and indicate basic opportunities of cost reduction via load shifting and energy storing (Turner et al., 2015). After the implementation of any energy conservation action through appropriate control systems, the manager recalculates the forecasts twice within the week to better calibrate and amend the existing plan. In order to apply such actions, the branch must be part of a larger scale electrical system and organized under a smart grid approach, while storing mechanisms must be ideally available (Favre and Peuportier, 2014).

In our experiments we implement four alternative forecasts. First, the methodology discussed above, see figure 3, that implements both decomposition and multiple temporal aggregation, through the MAPA framework. This will be named

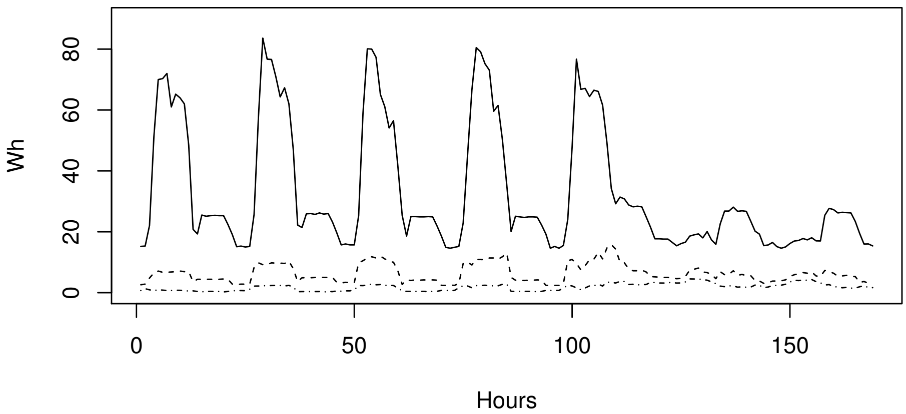

Fig. 5 Visualization of representative time series of the bank branches data set for a typical week: Energy consumption of all the branches (continuous), the first bank branch (dashed) and its HVAC use (dotted).

MAPA.D hereafter. Next, to evaluate the effect of MTA, we implement as a benchmark *SES*, after removing seasonality via decomposition, as in the methodology outlined for MAPA.D. To assess the impact of decomposition we implement the original *MAPA*, as described by Kourentzes et al. (2014), as well as the modified one, with the proposed weighting scheme described in section 3.1.2. The latter is named *MAPA.W*. We have considered an exponential smoothing base model with no decomposition and MTA, but has been excluded for brevity, as it did not perform well. As the results suggest the decomposition is particularly useful, due to the high dimensionality of the seasonal profile and the relatively limited sample size. The forecasting performance of the proposed methodology is evaluated both in term of forecasting accuracy (closeness of actual values and generated forecasts) and bias (consistent differences between actual values and generated forecasts). To this purpose we use the Relative Mean Absolute Error (RMAE) and Relative Absolute Mean Error (RAME):

$$RMAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{\sum_{i=1}^{n} |y_i - \hat{y_B}_i|},$$

$$RAME = \frac{\left|\sum_{i=1}^{n} y_i - \hat{y}_i\right|}{\left|\sum_{i=1}^{n} y_i - \hat{y_B}_i\right|},$$

where y_i are the actual values of series Y at point i, $\hat{y}_i$ the forecasts of the method being evaluated and $\hat{y_B}_i$ the forecasts of the method used as Benchmark. We summarize the metrics across time series using the geometric mean, resulting in ARMAE and ARAME for accuracy and bias. ARMAE has been proposed by (Davydenko and Fildes, 2013) (referred to as AvRelMAE by the authors) and ARAME is its bias equivalent. ARMAE has been shown to be robust to calculation issues, for instance overcoming limitations of the Geometric Relative Mean Absolute Error (GMRAE) that summarizes individual errors after the ratios are formed, and have minimal bias, in contrast to more popular metrics such as the Mean Absolute Percentage Errors (MAPE) or Mean Absolute Scaled Error (MASE, Hyndman and Koehler, 2006). Furthermore, the metric is easy to interpret. A result below one signifies an improvement over the benchmark forecast, while the opposite is true for values above 1. Percentage gains over the benchmark can be easily calculated as: $(1 - ARMAE)100\%$. We use SES as a benchmark in the calculation of the metrics.

Finally, we implement a rolling origin evaluation scheme (Tashman, 2000) to reduce the bias in our results. The original time series is divided into the training set, used to fit the model, and the test set, for evaluating its performance. Then, multiple evaluation rounds are performed as an additional observation is included into the fitting sample and updating the forecasting origin by one step at a time. Given a starting training set of length s and a forecasting horizon of h, a maximum number of $(n - s) - h + 1$ tests sets can be provided. We use the last 20% of observations as a test set, resulting in a two-weeks test set, providing a sample of 313 to 169 forecasts, depending on the forecasting horizon examined.

The analysis is performed using the R statistical software (R Core Team, 2018) and the packages of **MAPA**, which contains functions and wrappers for implementing the MAPA (Kourentzes and Petropoulos, 2018); **forecast**, which contains methods and tools for analysing time series (Hyndman et al., 2018); and **tsoutliers**, which contains functions for the detection of outliers in time series and their adjustment (de Lacalle, 2017).

5 Results

In tables 1 and 2 the performance of the cross-sectional aggregation methods is evaluated in terms of forecasting accuracy and bias for different forecasting horizons and for various hierarchical levels. In the first case, the performance is calculated by averaging the error metric values across the respective horizons (for short, medium and long-term) considering all levels, while in the latter by averaging the values across all the forecasting horizons and for each level separately. Note that in both tables SES is not reported as it is used as the denominator for the calculation of the metrics and the result is equal to 1 for every case.

Table 1 Accuracy (ARMAE) per forecasting horizon for the entire number of the series, and per hierarchical level across all the forecasting horizons tested.

Method	MAPA.D	MAPA	MAPA.W	MAPA.D	MAPA	MAPA.W
	All forecasting horizons and levels					
Bottom-up	0.861	1.433	0.978	0.861	1.433	0.978
Top-down	0.852	1.265	0.922	0.852	1.265	0.922
Optimal	**0.803**	1.333	0.917	**0.803**	1.333	0.917
	Short-term (t+1 to t+48)			*Level 0*		
Bottom-up	0.879	1.535	1.018	0.854	1.408	0.923
Top-down	0.859	1.322	0.939	**0.813**	1.333	0.897
Optimal	**0.814**	1.414	0.947	0.817	1.340	0.897
	Medium-term (t+49 to t+120)			*Level 1*		
Bottom-up	0.857	1.396	0.963	0.833	1.289	0.892
Top-down	0.847	1.237	0.912	0.857	1.270	0.916
Optimal	**0.797**	1.296	0.903	**0.802**	1.238	0.868
	Long-term (t+121 to t+168)			*Level 2*		
Bottom-up	0.848	1.372	0.953	0.899	1.620	1.136
Top-down	0.851	1.237	0.916	0.888	1.195	0.955
Optimal	**0.798**	1.292	0.903	**0.789**	1.427	0.991

Considering ARMAE, for the case of the MAPA.D, across all forecasting horizons (1-168) and levels, the optimal approach outperforms the rest of the hierarchical approaches. The same conclusion is made both for short, medium and long-term forecasts, as well as for predicting at the mid and bottom level of the hierarchy. At the top level, the top-down approach is marginally superior to the optimal. Similar results can be observed for ARAME.

In table 1 we can see that the benchmark SES is outperformed substantially by MAPA.D and MAPA.W, demonstrating the usefulness of MTA in modeling. MAPA.D that similarly to SES relies on decomposition, is overall superior to the non-decomposition based MAPA.W forecasts, by about 10%. The modified MAPA.W outperforms MAPA, as it caters for the high-frequency nature of the seasonality, but it is not more accurate than MAPA.D. This is attributed to the estimation challenges of the high-dimensional seasonal profile, with a relatively small sample size. MAPA.D avoids this estimation by employing decomposition. The same reasoning is applicable in explaining the relative poor performance of MAPA compared to SES (all ARMAE values are above 1).

Table 2 Bias (ARAME) per forecasting horizon for the entire number of the series, and per hierarchical level across all the forecasting horizons tested.

Method	MAPA.D	MAPA	MAPA.W	MAPA.D	MAPA	MAPA.W
	All forecasting horizons and levels					
Bottom-up	0.459	0.603	0.622	0.459	0.603	0.622
Top-down	0.488	0.527	0.744	0.488	0.527	0.744
Optimal	**0.412**	0.523	0.573	**0.412**	0.523	0.573
	Short-term (t+1 to t+48)			*Level 0*		
Bottom-up	0.686	0.820	0.791	0.399	0.421	0.596
Top-down	0.686	0.696	0.792	**0.357**	0.408	0.798
Optimal	**0.595**	0.675	0.692	0.403	0.412	0.759
	Medium-term (t+49 to t+120)			*Level 1*		
Bottom-up	0.416	0.554	0.547	0.449	0.476	0.618
Top-down	0.432	0.495	0.698	0.470	0.474	0.680
Optimal	**0.389**	0.489	0.523	**0.448**	0.488	0.635
	Long-term (t+121 to t+168)			*Level 2*		
Bottom-up	0.340	0.482	0.557	0.541	1.092	0.655
Top-down	0.392	0.426	0.744	0.691	0.758	0.759
Optimal	**0.302**	0.433	0.520	**0.387**	0.711	0.390

Considering the various hierarchical methods, we find that optimal combination performs overall best for most cases. For MAPA, which is poor at estimating the high-frequency seasonality compared to the alternative MAPA.D and MAPA.W, the top-down approach is beneficial, as it relies on estimation at the aggregate level, that the noise of the lower levels is not so strong. However, for the alternative forecasts that do not suffer from this limitation, the optimal combination allows to use information from all levels, resulting in the best accuracy.

Turning our attention to table 2 that provides the bias (ARAME) results, we observe similar findings. However, in this case all MAPA based forecasts are outperforming SES. Overall, the proposed MAPA.D outperforms all other alternatives, demonstrating the benefits of both MTA and decomposition. The optimal combination across hierarchical levels remains beneficial, as it allows using information from all levels of the hierarchy, in contrast to the bottom-up and top-down alternatives. However, in contrast to the accuracy results, the bottom-up approaches performs competitively to the top-down, echoing results in the literature that have found bottom-up to perform very well in terms of forecast bias (Athanasopoulos et al., 2009). Similarly, MAPA's bias is competitive to MAPA.D and MAPA.W, as the inaccurate modeling of seasonality is of less importance than the overall level of the forecasts in the calculation of the bias.

Regardless of the hierarchical reconciliation method used we find that both decomposition and MTA are beneficial, demonstrating the usefulness of the proposed approach. Reflecting on the differences between MAPA.D and MAPA.W, the former does not need to estimate the seasonal profile, reducing the optimization complexity. Furthermore, due to MTA it is robust against estimation uncertainty. MAPA.W gains both in terms of mitigating model uncertainty and parameter specification, evident in the superior results against the benchmark SES (both ARMAE and ARAME are consistently below 1), but due to the relatively limited sample size, it is not able to perform as well as MAPA.D. Another benefit of MTA is evident when comparing the differences in accuracy and bias between shorter and longer forecast horizons. Relatively to exponential smoothing, MAPA performs best at long forecast horizons. This finding is in agreement with the literature that argues this is due to the effect of incorporating information from the high-aggregation temporal levels, where long term dynamics are easier to model (Kourentzes et al., 2014).

Finally, we have experimented with MAPA forecasts that permit trend and found no substantial performance differences. We found that trend was rarely selected and in all cases it was strongly damped by MTA. The lack of strong trends was apparent in higher temporal aggregation levels, which in turn helped the final MAPA forecasts to have minimal trend. This again highlights the strength of MAPA in mitigating modeling uncertainty.

5.1 Implications for energy managers

The results of this study show that the proposed forecasting methodology can lead to significant improvements, especially when referring to long term forecasts of 6 to 7 days ahead. A key contribution of this work is the decision making support that the proposed methodology offers to energy managers. In order to optimize the energy use of the bank and its branches, detailed information is required regarding the energy intensive end uses of the buildings. The methodology provides such information across all hierarchical levels and enables the efficient monitoring and energy management of the system. In this regard, the energy manager can inspect the expected energy demand at the highest level of the hierarchy (bank), detect possible threats (problematic branches) and specify the cause of increased energy consumption (end uses). Energy optimization and conservation action plans, such as load shifting or maintenance of the facilities, will become easier to develop and implement and can become more targeted than present. Undoubtedly, reconciled forecasts is a prerequisite, which is a direct output of our modeling approach.

6 Conclusions

We proposed a holistic approach for forecasting effectively hierarchical electricity consumption time series, by producing both accurate and reconciled forecasts. This is key given that the forecasts of the lower aggregation levels of a system must always add up to the ones of the higher levels, and vice-versa. Multiple Temporal Aggregation is used, through the MAPA, to boost the forecasting performance and alleviate the effect of modeling uncertainty, while cross-sectional hierarchical approaches are applied to reconcile the individual forecasts across the hierarchy. Additionally, some modifications to MAPA's original form are introduced to enable it better capture the special characteristics of high-frequency data.

The results of our study show that temporal aggregation can lead to significant improvements in forecasting performance in terms of accuracy and bias, even when external variables which affect energy consumption are not considered and simple time series forecasting models like exponential smoothing are used instead. This is a promising outcome given that detailed regressor information is not always available, but also requires more complex forecasting models. When forecasting must be performed automatically for numerous time series, in order to support decisions within an acceptable time frame, fast, robust and reliable methods are required. We demonstrate that our proposed approach achieves this even for limited sample sizes. Furthermore, it can be easily implemented in existing forecasting support system, as it is based on exponential smoothing that is standard in most systems, offering a powerful alternative, where more complex methods mentioned in the literature, such as machine learning techniques, are not available or applicable.

Another finding of our study is that, apart from reconciling forecasts, cross-sectional aggregation can also enhance the forecasting performance by combining appropriately the base forecasts produced. We find that the optimal combination method that combines views of the time series from multiple levels of the hierarchy performs best, balancing the detailed information available at the bottom level, and the aggregate view of the higher levels.

Finally, although we find that the decomposition approach works best, we also find that the proposed weighting scheme for MAPA.W performs well against the original MAPA, avoiding the over-smoothing of the high-frequency seasonal profile. We attributed the better performance of decomposition to the relatively limited sample size. Nonetheless, it demonstrate that more research should be done in how the forecasting methods across the multiple temporal aggregation levels are combined, given the positive results of MTA reported here and in the literature.

References

Aslak Petersen J, Heide-Jørgensen DM, Detlefsen NK, Boomsma TK (2016) Short-term balancing of supply and demand in an electricity system: forecasting and scheduling. Annals of Operations Research 238(1):449–473

Athanasopoulos G, Ahmed RA, Hyndman RJ (2009) Hierarchical forecasts for australian domestic tourism. International Journal of Forecasting 25(1):146–166

Athanasopoulos G, Hyndman RJ, Kourentzes N, Petropoulos F (2017) Forecasting with temporal hierarchies. European Journal of Operational Research 262(1):60–74

Barrow D, Kourentzes N (2018) The impact of special days in call arrivals forecasting: A neural network approach to modelling special days. European Journal of Operational Research 264(3):967–977

Barzin R, Chen JJ, Young BR, Farid MM (2015) Peak load shifting with energy storage and price-based control system. Energy 92, Part 3:505–514

Beaumont AN (2014) Data transforms with exponential smoothing methods of forecasting. International Journal of Forecasting 30(4):918–927

Chalal ML, Benachir M, White M, Shrahily R (2016) Energy planning and forecasting approaches for supporting physical improvement strategies in the building sector: A review. Renewable and Sustainable Energy Reviews 64:761–776

Chen C, Liu LM (1993) Joint estimation of model parameters and outlier effects in time series. Journal of the American Statistical Association 88(421):284–297

Dangerfield BJ, Morris JS (1992) Top-down or bottom-up: Aggregate versus disaggregate extrapolations. International Journal of Forecasting 8(2):233–241

D'Attilio DF (1989) Practical applications of trend analysis in business forecasting. The Journal of Business Forecasting Methods & Systems 8:9–11

Davydenko A, Fildes R (2013) Measuring forecasting accuracy: The case of judgmental adjustments to sku-level demand forecasts. International Journal of Forecasting 29(3):510–522

Dudek G (2016) Pattern-based local linear regression models for short-term load forecasting. Electric Power Systems Research 130:139–147

Erişen E, Iyigun C, Tanrısever F (2017) Short-term electricity load forecasting with special days: an analysis on parametric and non-parametric methods. Annals of Operations Research

Favre B, Peuportier B (2014) Application of dynamic programming to study load shifting in buildings. Energy and Buildings 82:57–64

Fliedner EB, Lawrence B (1995) Forecasting system parent group formation: An empirical application of cluster analysis. Journal of Operations Management 12(2):119–130

Gardner ES (1985) Exponential smoothing: the state of the art. Journal of Forecasting 4(1):1–28

Gordon T, Morris J, Dangerfield B (2000) Top-down or bottom-up: which is the best approach to forecasting? The Journal of Business Forecasting Methods & Systems 16(3):13–16

Gross CW, Sohl JE (1990) Disaggregation methods to expedite product line forecasting. Journal of Forecasting 9(3):233–254

Handik W, S V, Rajesh P (2007) On the effectiveness of top-down strategy for forecasting autoregressive demands. Naval Research Logistics 54(2):176–188

Heiple S, Sailor DJ (2008) Using building energy simulation and geospatial modeling techniques to determine high resolution building sector energy consumption profiles. Energy and Buildings 40(8):1426–1436

Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, Petropoulos F, Razbash S, Wang E, Yasmeen F (2018) forecast: Forecasting functions for time series and linear models. URL http://pkg.robjhyndman.com/forecast, r package version 8.4

Hyndman RJ, Koehler AB (2006) Another look at measures of forecast accuracy. International Journal of Forecasting 22(4):679–688

Hyndman RJ, Koehler AB, Snyder RD, Grose S (2002) A state space framework for automatic forecasting using exponential smoothing methods. International Journal of Forecasting 18(3):439–454

Hyndman RJ, Ahmed RA, Athanasopoulos G, Shang HL (2011) Optimal combination forecasts for hierarchical time series. Computational Statistics & Data Analysis 55(9):2579–2589

Jurado S, Nebot À, Mugica F, Avellana N (2015) Hybrid methodologies for electricity load forecasting: Entropy-based feature selection with machine learning and soft computing techniques. Energy 86:276–291

Kavgic M, Mavrogianni A, Mumovic D, Summerfield A, Stevanovic Z, Djurovic-Petrovic M (2010) A review of bottom-up building stock models for energy consumption in the residential sector. Building and Environment 45(7):1683–

1697
Kendall M, Stuart A (1983) The advanced theory of statistics. Griffin 3:410–414
Kohn R (1982) When is an aggregate of a time series efficiently forecast by its past? Journal of Econometrics 18(3):337–349
Kourentzes N, Petropoulos F (2016) Forecasting with multivariate temporal aggregation: The case of promotional modelling. International Journal of Production Economics 181, Part A:145–153, DOI http://dx.doi.org/10.1016/j.ijpe.2015.09.011
Kourentzes N, Petropoulos F (2018) MAPA: Multiple Aggregation Prediction Algorithm. URL https://CRAN.R-project.org/package=MAPA, r package version 2.0.4
Kourentzes N, Petropoulos F, Trapero JR (2014) Improving forecasting by estimating time series structural components across multiple frequencies. International Journal of Forecasting 30(2):291–302
Kourentzes N, Rostami-Tabar B, Barrow DK (2017) Demand forecasting by temporal aggregation: using optimal or multiple aggregation levels? Journal of Business Research 78:1–9
de Lacalle JL (2017) tsoutliers: Detection of Outliers in Time Series. URL https://CRAN.R-project.org/package=tsoutliers, r package version 0.6-6
Lai SH, Hong T (2013) When One Size No Longer Fits All: Electric Load Forecasting with a Geographic Hierarchy. URL http://assets.fiercemarkets.net/public/sites/energy/reports/electricloadforecasting.pdf, sAS White Paper
Ledolter J (1989) The effect of additive outliers on the forecasts from arima models. International Journal of Forecasting 5(2):231–240
Miller DM, Williams D (2003) Shrinkage estimators of time series seasonal factors and their effect on forecasting accuracy. International Journal of Forecasting 19(4):669–684
Nikolopoulos K, Syntetos AA, Boylan JE, 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
Pedregal DJ, Trapero JR (2010) Mid-term hourly electricity forecasting based on a multi-rate approach. Energy Conversion and Management 51(1):105–111
Petropoulos F, Kourentzes N (2014) Improving forecasting via multiple temporal aggregation. Foresight: The International Journal of Applied Forecasting 2014(34):12–17
Petropoulos F, Kourentzes N (2015) Forecast combinations for intermittent demand. Journal of the Operational Research Society 66(6):914924
R Core Team (2018) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/
Schwarzkopf AB, Tersine RJ, Morris JS (1988) Top-down versus bottom-up forecasting strategies. International Journal of Production Research 26(11):1833–1843
Shlifer E, Wolff R (1979) Aggregation and proration in forecasting. Management Science 25(6):594–603
Silvestrini A, Veredas D (2008) Temporal aggregation of univariate and multivariate time series models: a survey. Journal of Economic Surveys 22(3):458–497

Spithourakis G, Petropoulos F, Babai MZ, 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

Suganthi L, Samuel AA (2012) Energy models for demand forecastinga review. Renewable and Sustainable Energy Reviews 16(2):1223–1240

Tashman LJ (2000) Out-of-sample tests of forecasting accuracy: an analysis and review. International Journal of Forecasting 16(4):437–450

Tiao G, Guttman I (1980) Forecasting contemporal aggregates of multiple time series. Journal of Econometrics 12(2):219–230

Trapero JR, Kourentzes N, Martin A (2015) Short-term solar irradiation forecasting based on dynamic harmonic regression. Energy 84:289–295

Tratar LF, Strmčnik E (2016) The comparison of holtwinters method and multiple regression method: A case study. Energy 109:266–276

Turner W, Walker I, Roux J (2015) Peak load reductions: Electric load shifting with mechanical pre-cooling of residential buildings with low thermal mass. Energy 82:1057–1067

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

Widiarta H, Viswanathan S, Piplani R (2008) Forecasting item-level demands: an analytical evaluation of topdown versus bottomup forecasting in a production-planning framework. IMA Journal of Management Mathematics 19(2):207–218

Zhao HX, Magoulès F (2012) A review on the prediction of building energy consumption. Renewable and Sustainable Energy Reviews 16(6):3586–3592

Zotteri G, Kalchschmidt M, Caniato F (2005) The impact of aggregation level on forecasting performance. International Journal of Production Economics 93 94:479–491