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

Evangelos Spiliotis[a,*], Fotios Petropoulos[b], Nikolaos Kourentzes[c], Vassilios Assimakopoulos[a]

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

## Abstract

Achieving high accuracy in energy consumption forecasting is critical for improving energy management and planning. However, this requires the selection of appropriate forecasting models, able to capture the individual characteristics of the series to be predicted, which is a task that involves a lot of uncertainty. When hierarchies of load from different sources are considered together, the uncertainty and complexity increase 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 forecasting, such as the bottom-up, the top-down, and the optimal reconciliation methods, can address the aggregation consistency concerns, it does not resolve the model selection uncertainty. To address this issue, we rely on Multiple Temporal Aggregation (MTA), which has been shown to mitigate the model selection problem for low-frequency time series. We propose a modification of the Multiple Aggregation Prediction Algorithm, a special implementation of MTA, for high-frequency time series to better handle the undesirable effect of seasonality shrinkage that MTA implies and combine it with conventional cross-sectional hierarchical forecasting. The impact of incorporating temporal aggregation in hierarchical forecasting is empirically assessed using a real data set from five bank branches. We show that the proposed MTA approach, combined with the optimal reconciliation method, demonstrates superior accuracy, aggregation consistency, and reliable automatic forecasting.

*Keywords:* Temporal aggregation, Hierarchical forecasting, Electricity consumption, Exponential smoothing, Seasonality shrinkage

## Nomenclature

**Abbreviations**

MAPA    Multiple Aggregation Prediction Algorithm

MTA      Multiple Temporal Aggregation

SES      Simple Exponential Smoothing

**Variables**

$\alpha$         smoothing parameter of SES

$b_i$        the trend component of the series at point $i$

$\bar{b}_h$       the median of the trend component of the series estimated for the h-step-ahead forecast produced across all temporal aggregation levels

$e_t$        forecast error at time $t$

$f$         time series periodicity

$h$         forecasting horizon

$k$         hierarchical level

$K$        the lowest level of the hierarchy

$l_i$        the level component of the series at point $i$

$\bar{l}_h$       the median of the level component of the series estimated for the h-step-ahead forecast produced across all temporal aggregation levels

$m$       total number of series in the hierarchy

$m_i$      total number of series at hierarchical level $i$

$n$        number of historical observations (length of series)

$p_j$       average of the historical proportions of the j bottom level series relative to the total aggregate $Y_t$

$s_i$       the seasonality component of the series at point $i$

$^*$Corresponding author

*Email addresses:* `spiliotis@fsu.gr` (Evangelos Spiliotis), `f.petropoulos@bath.ac.uk` (Fotios Petropoulos), `n.kourentzes@lancaster.ac.uk` (Nikolaos Kourentzes), `vassim@fsu.gr` (Vassilios Assimakopoulos)

| | |
|---|---|
| $\bar{s}_h$ | the median of the seasonality component of the series estimated for the h-step-ahead forecast produced across all temporal aggregation levels |
| $t$ | time period |
| $y_t$ | observation of series Y at time $t$ |
| $\hat{y}_t$ | forecast of series Y at time $t$ |
| $Y_t$ | aggregate of all series at time $t$ |
| $Y_{x,t}$ | the $x^{th}$ series at time $t$ used to disaggregate series $Y_t$ |
| $\hat{Y}_{x,n}(\text{h})$ | h-step-ahead (base) forecasts of series $x$ |
| $Y^{[k]}$ | temporally aggregated time series at level k |

**Vectors and matrices**

| | |
|---|---|
| $\mathbf{I}_k$ | identity matrix of order $k \times k$ |
| $\mathbf{0}_{l \times k}$ | null matrix of order $l \times k$ |
| $\mathbf{p}$ | vector consisting of all the historical proportions of the hierarchy |
| $\mathbf{P}$ | a matrix of order $m_K \times m$ used to reconcile the base forecasts |
| $\mathbf{S}$ | summing matrix representing the hierarchical structure |
| $\mathbf{Y_{i,t}}$ | vector consisting of all the observations of level $i$ at time $t$ |
| $\hat{\mathbf{Y}}_n(h)$ | vector consisting of the h-step-ahead (base) forecasts |
| $\tilde{\mathbf{Y}}_n(h)$ | vector consisting of the final h-step-ahead (reconciled) forecasts |

## 1. Introduction

Energy consumption forecasting encompasses a wide range of forecasting problems. Achieving high forecast accuracy can yield significant improvements to energy management, leading to many economic and environmental benefits through energy conservation techniques such as load shifting, peak shaving, and energy-storing [1, 2, 3]. Moreover, it is crucial for maintaining the balance between the load and generation [4], as well as for planning the expansion of power grids [5]. Motivated by these potential gains, substantial work has been done to improve energy forecasting methods and models, including statistical [6, 7], machine learning [8], deep learning [9, 10, 11], and

hybrid [12, 13] approaches among others. Additionally, different methods exist according to the forecasting horizon considered, given that different decisions are supported per case [14].

The literature focuses on three main classes of methods based on the prediction models used: statistical, engineering, and artificial intelligence methods. Suganthi and Samuel [15] provide a review of these methods, while Zhao and Magoulès [16] focus on those used in the building sector. They find that each forecasting method has its strengths and weaknesses concerning the problem at hand, available data, and the level of acceptable complexity. The literature does not identify the best method, and therefore, the model selection problem remains an unresolved key modeling issue [17]. This is particularly relevant in practice, where a reliable selection of forecasts is desirable.

To mitigate this problem, in the general time series forecasting context, Kourentzes et al. [18] proposed using Multiple Temporal Aggregation (MTA), later generalized by [19]. 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 [20]. Mainstream time series modeling literature has mainly focused on identifying the single optimal level that makes modeling simpler [21]. 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 to energy conservation, management, and pricing, any decision taken is multi-layered, considering and affecting multiple levels of the energy system it refers to [22, 23]. For example, when optimizing the 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. On the one hand, this requires model selection and estimation for many time series and 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 hierar-

chical forecasting methods, such as "bottom-up" and "top-down", have been proposed to achieve reconciliation [24].

Given the need for applying cross-sectional approaches to the problem at hand, the question arises whether 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 is realized using the Multiple Aggregation Prediction Algorithm (MAPA) by Kourentzes et al. [18]. We propose a modification to make MAPA appropriate for high-frequency time series, i.e., better handling the undesirable effect of seasonality shrinkage that typical MTA approaches imply [25], and introduce an approach to combine conventional cross-sectional hierarchical forecasting with MAPA. Our results show that our approach contributes towards *(a)* decreasing model uncertainty and increasing accuracy while *(b)* ensuring reconciled forecasts across the hierarchy. Both enable automation of forecasting in such problems, aiding decision-makers. Based on the above, our contribution to the literature is twofold: *(i)* we evaluate the beneficial effect of cross-temporal aggregation in electricity consumption forecasting, and *(ii)* we provide an approach for effectively applying MTA to seasonal time series.

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 presented in section 4. Section 5 summarizes the results, followed by concluding remarks in section 6.

## 2. Literature Review

### 2.1. Cross-sectional hierarchies for forecasting

Energy applications are closely related to 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 for exploiting the information hidden in hierarchies.

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 inves-

tigated a variety of cross-sectional hierarchical approaches that can produce a reconciled forecast [26]. 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 [27]. 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. [24] 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 [24, 19, 28, 29]. The top-down approach is considered to be more appropriate for long-term forecasts [30], as it effectively captures the trend of the data [31]. On the other hand, the bottom-up approach performs better among highly correlated time series [32] as it highlights the unique characteristics of the disaggregated data [33, 30], 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 [34, 35]. The correlation of the individual time series and their errors [36], as well as their variability [37, 38] might indicate which approach is preferable. However, the performance differences between the two approaches can often be minimal in practice, also displaying insignificant advantages over other formal or informal strategies reported in the literature [39]. For instance, Widiarta et al. [40] used both the top-down and the bottom-up methods to predict the demand of various items, concluding that their performance is similar, especially when the autocorrelation of the demand data is low [41]. Information from all hierarchical levels could be considered instead, with evidence of benefits for the overall forecasting performance [26, 24, 29].

In the field of energy consumption forecasting, both top-down and bottom-up approaches are used for energy planning and management [42, 43]. However, the latter is more popular [44], given that energy models usually correlate consumption with temperature data, which are monitored at the lowest levels of the hierarchy. More recently, Lai and Hong [45] 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 samples may obscure the underlying structure of the observed series and affect parameter estimation [46]. This, in turn, can change the identified model structure, even if we assume that the appropriate model family is chosen, which itself is uncertain [16]. 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 the temporal aggregation of time series. Silvestrini and Veredas [21] 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 there are merits in using temporal aggregation, but concluded it is difficult to identify the optimal temporal aggregation level. Weiss [47] offered insights into 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" [48, 49] by revealing patterns which are more evident in lower frequencies. Yet, the difficulty in identifying the optimal aggregation level and selecting an appropriate model remains an issue [50].

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 [51]. 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 [18]. This is particularly relevant to fast-moving data, such as electricity consumption forecasting applications, where the high sampling frequency displays increased noise and introduces multiple seasonal patterns at the original sampling frequency, which require data pre-processing of high complexity [52].

MTA was proposed by Kourentzes et al. [18] as implemented in MAPA, although the term itself was coined by Petropoulos and Kourentzes [51] as a more general concept. Since different

variations of MTA have appeared, notably the Temporal Hierarchies [19] and specialized variants for intermittent demand [50] and promotional modeling [53]. MAPA models multiple aggregated views of a time series, using independent exponential smoothing models. The resulting outputs of the models from each aggregation level are then combined to produce a final forecast. The key advantage of the approach is that by using a different model per frequency, various time series components are captured, as these are differentially highlighted in different temporal aggregation levels. Moreover, 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 [53, 54]. Kourentzes et al. [55] 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 [56]. In this respect, we investigate how to effectively deal with model uncertainty in complex energy consumption hierarchies of high-frequency data, particularly in consumption forecasting applications, while maintaining good forecasting performance. The proposed methodological approach combines these commonly separate aggregation frameworks, cross-sectional and temporal, to gain reconciled forecasts of reduced modeling complexity while putting lots of emphasis on the optimal handling of the seasonal component of the energy data, which is typically shrunk by conventional MTA implementations [25]. As far to our knowledge, the only studies considering such a combination, but still evaluated in different applications than electricity consumption, are those of Kourentzes and Athanasopoulos [57], generating coherent monthly forecasts for Australia tourism flows across both cross-sections and planning horizons, and Yagli et al. [58], applying spatial-temporal reconciliation for generating day-ahead forecasts for photovoltaic power generation plants in California.

## 3. Methodology

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

### 3.1. Aggregation and forecasting methods

The cross-sectional and temporal aggregation will be combined within the framework 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 $K$ to 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 $\mathbf{Y}_t = [Y_t, \mathbf{Y}_{1,t}, \ldots, \mathbf{Y}_{K,t}]$. 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 un-reconciled base forecasts. Using the notation above, $\tilde{\mathbf{Y}}_n(h) = \mathbf{S}\mathbf{P}\hat{\mathbf{Y}}_n(h)$, where $\mathbf{P}$ is an appropriate matrix of order $m_K \times m$ and $\tilde{\mathbf{Y_n}}(\mathbf{h})$ 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 $\mathbf{I}_k$

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 $\mathbf{p} = [p_1, p_2, \ldots, p_{m_K}]$ 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

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

where $p_j$ reflects the average of the historical proportions of the $j = 1, \ldots, k$ bottom level series $Y_{j,t}$ over the period $t = 1, \ldots, n$ 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. [26].

Finally, the *optimal combination* identifies $\mathbf{P}$ to provide minimal reconciliation errors, i.e., enforce aggregation consistency across forecasts by requiring only minimal changes of the base forecasts. Hyndman et al. [24] 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, which, on the other hand, requires reasonable base 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. It may depend on the type of data, the application, and the 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 drawing 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 to combine the forecasts, all 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 into 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 an aggregation level a trend is identified, but at another none is identified and set to zero, we do not indicate to prefer one or the other option. Therefore, these are combined

11

into a damped trend. Naturally, if most levels identify no trends, then any estimated trend will be diminished and vice-versa.

In forecasting electricity consumption data, there is a crucial consideration that should not be overlooked: accurate prediction of peaks is important [59]. Peak load 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 [25]. 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 consumption 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 peaks 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-seasonalize the final forecasts. A deterministic seasonality is forced, helping us to handle the peaks effectively. 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.

This approach makes the use of 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) [60]. In this regard, the final forecast of MAPA will be the median of the levels calculated. To support this simplification, Figure 2 presents the forecasts produced by seasonally adjusted MAPA for the same series examined in Figure 1, but this time by allowing the estimation of the trend. As seen, MAPA
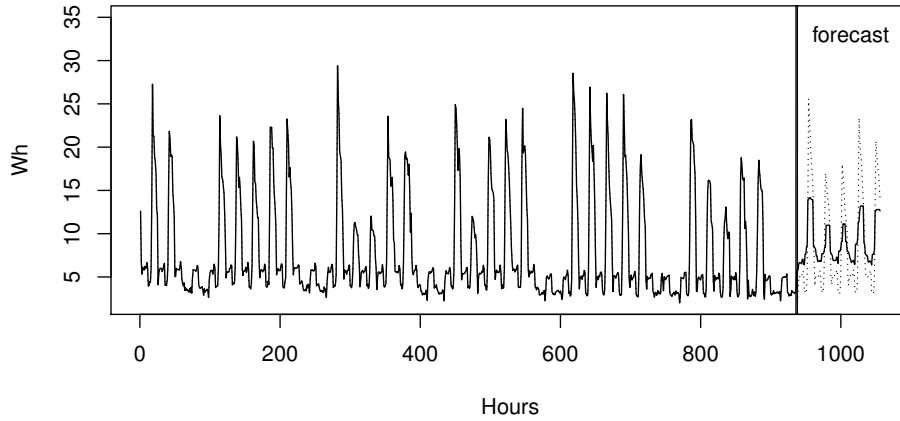
12

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

does not identify any significant trends across the various temporal levels considered, resulting in identical forecasts with that of the simplified approach.
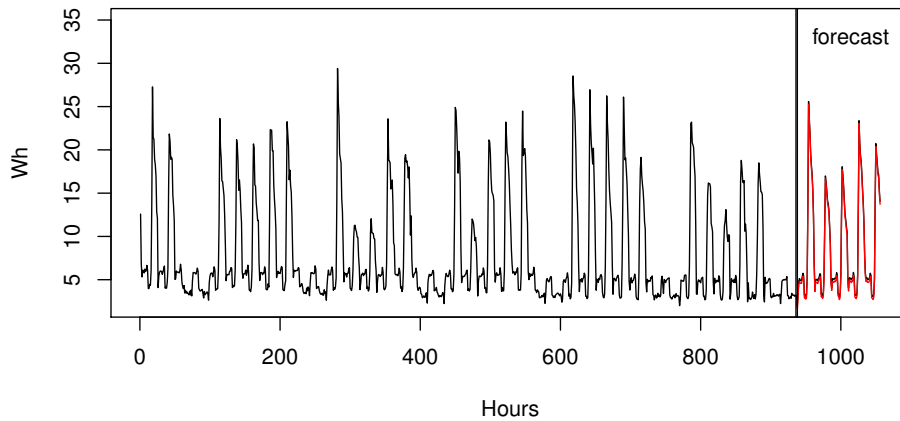


Figure 2: The effect of seasonally adjusted MAPA to hourly electricity consumption of a commercial building with strong seasonality when the estimation of trend is either allowed (red) or not (black). Due to the lack of significant trends, the two approaches result in identical results.

Undoubtedly, if the same forecasting method is to be applied to all temporally aggregated views

of the 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 of the series. SES produces forecasts using a single estimation of the smoothing parameter and initial level state. Both of those parameters are specified through the appropriate criteria. However, there is always the risk of inadequate parameterization due to the effect of outliers and other unusual values, especially for a 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 of both level and trend components remains unweighted. Kourentzes et al. [18] 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 is not desirable, as it is beneficial [61]. Once again, we use the time series of Figure 1 to justify our claims. As shown in Figure 3, seasonally weighted MAPA does mitigate the effect of seasonal shrinkage, providing results similar to those of seasonally adjusted MAPA. Yet, it seems that this approach still underestimates the peak load.

The decomposition approach simplifies the specification of MAPA substantially, considering both the number of parameters to be estimated (in exponential smoothing the highest 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
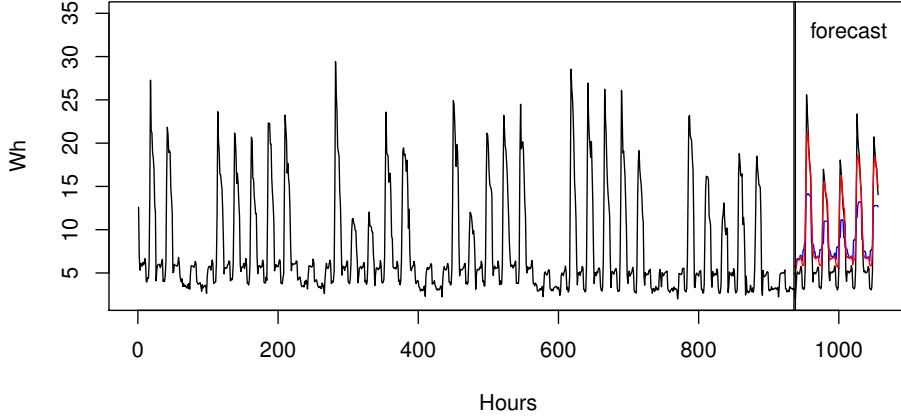
Figure 3: The effect of MAPA (blue) on hourly electricity consumption of a commercial building with strong sea-sonality. Both seasonally adjusted MAPA (black) and seasonally weighted MAPA (red) provide less shrunk seasonal indexes than the original implementation. However, seasonally weighted MAPA still underestimates the peak load.

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 initially proposed form, leading potentially to improvements in forecasting performance.

### 3.1.3. Exponential smoothing

The Simple 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:

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

15

where $l_t$ is the estimated level of the series and $\hat{y}_t$ the forecast of SES at point $t$. $\alpha$ 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 $\alpha$, 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 $\alpha$. This is done by maximising the likelihood $\mathcal{L}$ of the model [62]:

$$\mathcal{L}(\alpha, l_0) = -\frac{n}{2} log(\sum_{t=1}^{n} (e_t)^2),$$

where $n$ is the length of the series and the error $e_t$ is conditional on the smoothing parameter $\alpha$ 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 is true for the case of the trend component. All typical variants of exponential smoothing are described by Hyndman et al. [62]. 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 decrease 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 [63]. Therefore, data cleansing becomes a task of significant importance [17].

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 essential data consideration is special days, such as bank holidays, which can affect the forecasting performance negatively [64]. These can reduce accuracy during both outlying and regular periods. Barrow and Kourentzes [54] 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 [65]. 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 using classical decomposition by moving averages [66], with a seasonal periodicity of 168 hours. We use additive decomposition, 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. 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 4). 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 profiles and vice versa. Observe that the weekly profile has substantially lower variation than the daily one, indicating that the former is estimated more accurately and is preferable to the daily one.
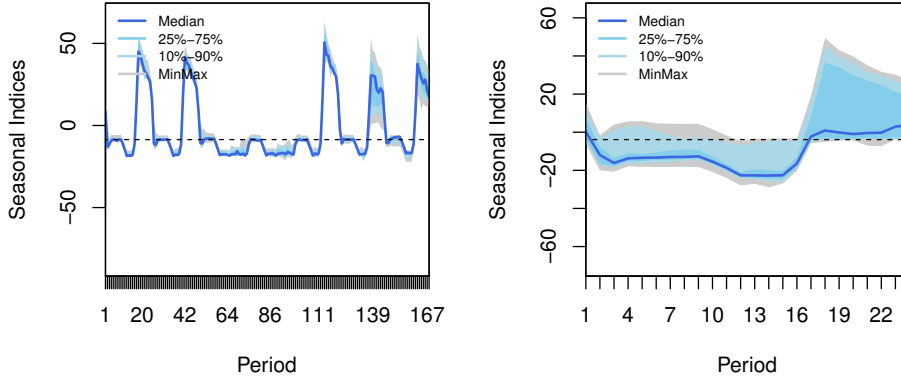
17

Figure 4: 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 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 [67] 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-seasonalized, 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 which cross-sectional aggregation approach is the best, we retain all and evaluate the best one. An overview of the proposed methodology is presented in Figure 5.

## 4. Experimental Design

### 4.1. Data and case study

The proposed methodology is applied to a group of five bank branches located in Athens, Greece. We examine the benefits in terms of accuracy, complexity, and decision support.
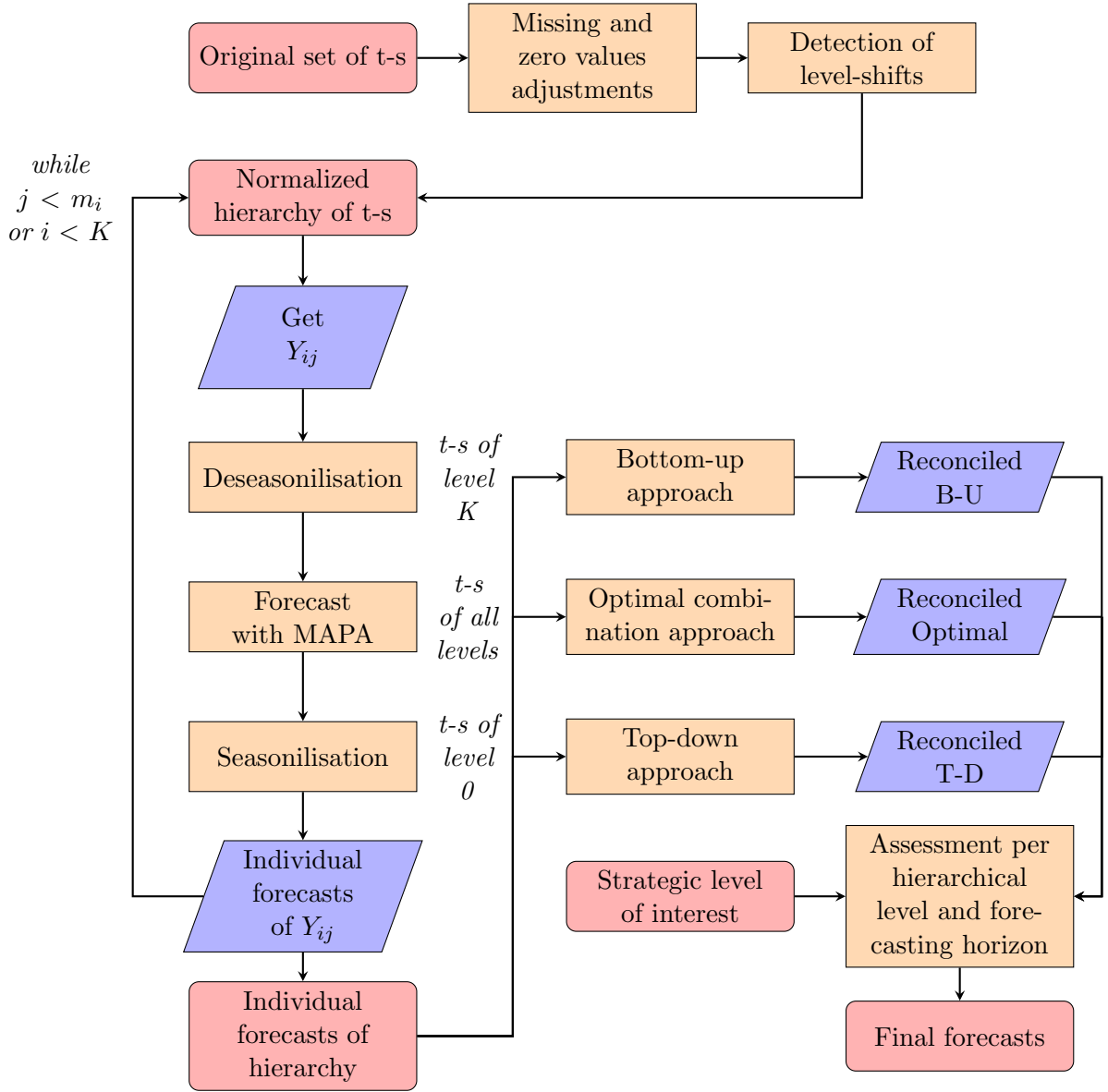
18

Figure 5: Flow chart of the proposed methodological framework applied to a K-level hierarchy. $Y_{ij}$ indicates the $j_{th}$ time series of level $i$.

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 6, while a typical example of the time series of each level is provided in Figure 7. 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 approximately 3%. The majority of them belong in the training 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 [8]. 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.
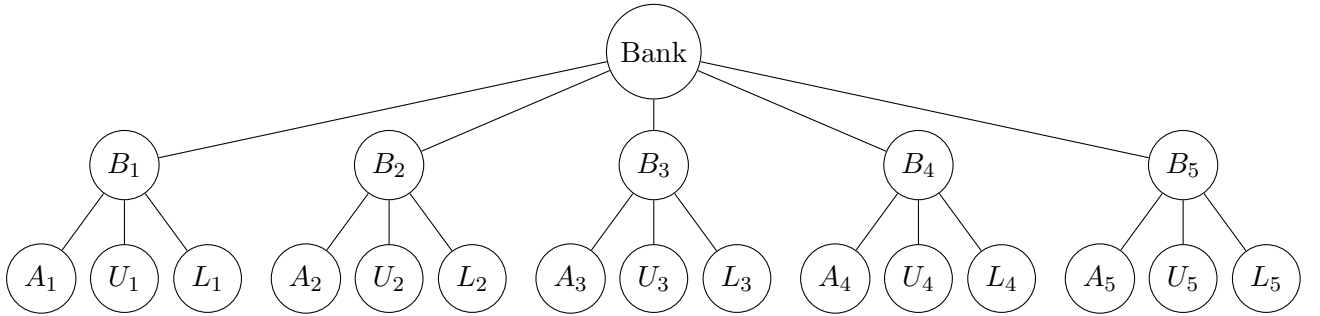


Figure 6: The three-level hierarchical tree diagram of the bank case-study. $B_i$, $A_i$, $U_i$ and $L_i$ stand for the $i_{th}$ **B**ranch, HV**A**C, **U**PS, and **L**ighting energy use, respectively.

### 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 relevant decisions. More specifically, we examine three windows that mirror the current bank's energy manager practices: up to 2 days (1-48 hours), up to 5 days (49-120 hours), and up to a week. Thus, the most appropriate combination of temporal and cross-sectional aggregation methods will be empirically demonstrated.

At the beginning of every week, forecasts are produced for all branches to highlight possible threats and indicate necessary opportunities for cost reduction via load shifting and energy-storing [68]. After the implementation of any energy conservation action through appropriate control
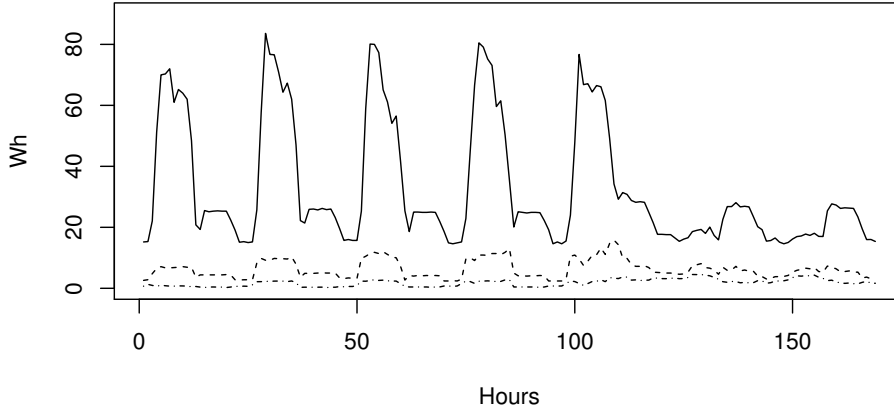
Figure 7: Visualization of a representative time series of the bank branches data set for a typical week: Energy consumption of all the branches (continuous), the first bank branch (dashed), and its HVAC use (dotted).

systems, the manager recalculates the forecasts twice within the week to better calibrate and amend the existing plan. To apply such measures, 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 [69].

In our experiments, we implement four alternative forecasts. First, we consider the methodology discussed above (see Figure 5), which implements both decomposition and multiple temporal aggregation, through the MAPA framework. This will be named *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 and seasonal adjustment, we apply the original *MAPA*, as described by [18], 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 also tested an exponential smoothing base model with no decomposition and MTA but we do not present its performance 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 terms of forecasting accuracy (closeness of actual values and generated forecasts) and bias (consistent differences

21

between actual values and generated forecasts). To this purpose, we use the Relative Mean Absolute Error (RMAE) and Relative Absolute Mean Error (RAME):

$$RMAE = \frac{\sum_{i=n+1}^{h} |y_i - \hat{y}_i|}{\sum_{i=1}^{n} |y_i - \hat{y}_{B_i}|},$$

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

where $y_i$ are the actual values of series $Y$ at point $i$, $\hat{y}_i$ the forecasts of the method being evaluated, $\hat{y}_{B_i}$ the forecasts of the method used as $Benchmark$, and $h$ the forecasting horizon. 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 [70] (referred to as AvRelMAE by the authors), and ARAME is its bias equivalent. ARMAE has been shown to be robust to calculation issues, overcoming limitations of the Geometric Relative Mean Absolute Error (GMRAE) that summarizes individual errors after the ratios are formed. ARMAE also has a minimal bias, in contrast to more popular metrics such as the Mean Absolute Percentage Errors (MAPE) [71]. Furthermore, the metric is easy to interpret. A value 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 [72] 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 in the fitting sample and updating the forecasting origin by one step at a time. Given an initial training set of length $s$ and a forecasting horizon of $h$, a maximum number of $(n - s) - h + 1$ validation 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 [73] and the packages of **MAPA**,

which contains functions and wrappers for implementing the MAPA [74]; **forecast**, which contains methods and tools for analysing time series [75]; and **tsoutliers**, which contains functions for the detection of outliers in time series and their adjustment [76].

## 5. Results

In tables 1 and 2, we present the performance of the cross-sectional aggregation methods in terms of forecasting accuracy and bias for different forecasting horizons and various hierarchical levels. In the first case, the performance is calculated by averaging the error measure values across the respective horizons (for all horizons) 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.

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 all the forecasting horizons considered, 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 to 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 relatively small sample size. MAPA.D avoids this estimation by employing decomposition. The same reasoning is applicable in explaining the relatively poor performance of MAPA compared to SES (all ARMAE values are above 1).

Considering the various hierarchical methods, we find that optimal combination performs overall best for most cases. For MAPA, which is mediocre 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. How-

23

Table 1: Accuracy (ARMAE) per forecasting horizon per hierarchical level across and forecasting horizons.

| 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 |
| | *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 |
| | *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 |
| | *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 |

Table 2: Bias (ARAME) per forecasting horizon per hierarchical level across and forecasting horizons.

| 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 |
| | *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 |
| | *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 |
| | *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 |

ever, for the alternative forecasts that do not suffer from this limitation, the optimal combination allows using 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 perform competitively to the top-down, echoing findings in the literature that have found bottom-up to perform very well in terms of forecast bias [26]. 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 longer 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 more natural to model [18].

Finally, we have experimented with MAPA forecasts that permit trends and found no substantial performance differences. We discovered that a trend component 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 a 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 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. To optimize the energy use of a building system and its components, detailed information is required regarding the energy-intensive end uses of the individual buildings [77]. 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 the present. Undoubtedly, reconciled forecasts is a prerequisite, which is a direct output of our modeling approach.

Given its generalized nature and reasonable complexity, the proposed methodology can be easily implemented in any block of buildings, such as retail stores and shopping centers, public buildings, bank branches, offices, hotels, and cinemas. Our methodology could give support to Intelligent Energy Management Systems (IEMS) that assist energy managers grand better control and monitoring, prevent costs and contamination, as well as ensure comfort and wellness [8]. The beneficial effects of the methodology could become even more significant if connected to Smart Energy Management Systems (SEMS), utilizing smart meters to optimize appliance scheduling in a $h$-hours ahead period and allocate energy resources of appliances in real-time [78]. Finally, the forecasts provided by the methodology could serve as benchmarks for detecting energy efficiency anomalies in smart buildings [79] and minimizing the risk of malfunctions and deterioration [80].

### 5.2. On the effect of seasonality shrinkage

One of the main contributions of this study is the introduction of an easy to implement methodology that allows the generation of robust forecasts through MTA, while also accurately capturing the seasonal component of energy consumption series that are characterized by strong periodic fluctuations and peaks. This is a fundamental concept in energy consumption forecasting, closely related to most energy-saving, efficiency, and conservation actions. However, the proposed methodology could be exploited to improve forecasting accuracy in almost any application involving the

extrapolation of high-frequency data, mitigating the effect of seasonality damping that typical MTA approaches imply.

To demonstrate the value added by the proposed methodology in such applications, we consider the 1428 monthly and 756 quarterly series of the M3 competition [81], the standard testing ground of generic time series forecasting algorithms. Like in Section 4.2, we consider three different implementations of MTA: (i) MAPA, implementing the original MAPA framework, (ii) MAPA.D, implementing both decomposition and MTA, and (iii) MAPA.W, implementing the weighted combination scheme of MAPA for retaining seasonality. Moreover, since M3 involves business data of various domains (micro, industry, macro, finance, and demographic) that are characterized by diverse features, such as trend, seasonality, and auto-correlation [82], instead of considering just SES, we allow MAPA to consider any possible model of the ExponenTial Smoothing family (ETS), as described by Hyndman et al. [62]. Thus, in contrast to the previous case-study, $\bar{b}_h$ of (2) can be different than zero.

Table 3 summarizes the performance of the examined approaches, both per frequency and in total. We use ARMAE and ARAME for measuring forecasting accuracy and bias, using ETS instead of SES as a benchmark in the calculation of the metrics to enable fair comparisons. As seen, all implementations display metric values lower than one, highlighting the benefits of applying MTA. However, MAPA.D and MAPA.W outperform MAPA both in terms of accuracy and bias, indicating that seasonality shrinkage of MTA does have a significant effect on forecasting performance. Moreover, in most cases, MAPA.D performs better than MAPA.W, especially for the quarterly series, where the sample size is more limited. In this regard, we conclude that MTA with seasonal decomposition is a promising alternative to standard MTA that should be considered when extrapolating seasonal series.

## 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.

In our approach, Multiple Temporal Aggregation (MTA) is used, through the Multiple Ag-

28

Table 3: Accuracy (ARMAE) and bias (ARAME) for the 756 quarterly and 1,428 monthly series of the M3 Competition. The performance of the proposed methodology (*MAPA.D*) is compared to that of the original *MAPA* as well as its weighed combination scheme (*MAPA.W*) to identify best practices for applying MTA when dealing with seasonal data.

| Dataset | MAPA.D | MAPA | MAPA.W |
|---|---|---|---|
| | *Accuracy* | | |
| Quarterly | **0.915** | 0.949 | 0.931 |
| Monthly | **0.918** | 0.941 | 0.919 |
| Total | **0.917** | 0.944 | 0.923 |
| | *Bias* | | |
| Quarterly | **0.934** | 0.976 | 0.976 |
| Monthly | 0.979 | 0.981 | **0.965** |
| Total | **0.963** | 0.979 | 0.969 |

gregation Prediction Algorithm (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 to capture better the unique characteristics of high-frequency data and deal with seasonality shrinkage that typical MTA approaches imply. The results of our study indicate that:

- MTA significantly improves forecasting performance in terms of accuracy and bias.

- Cross-sectional aggregation further enhances forecasting performance by combining appropriately the base forecasts produced.

- Applying MTA to seasonally adjusted data leads to better forecasts than applying MTA to the original series.

More specifically, we find that the optimal combination method, which combines views of the time series from multiple levels of the hierarchy, performs best when combined with MTA. Thus, we confirm that balancing the detailed information available at the bottom level of the hierarchy and the aggregate view of its higher levels is the best strategy for improving forecasting performance.

We also find that weighting appropriately the seasonal components computed by MAPA across different temporal levels (MAPA.W) leads to better forecasts than the original MAPA. Thus, we confirm the benefits of avoiding the over-smoothing of the high-frequency seasonal profile. Furthermore, we attribute the better performance of the proposed approach (MAPA.D) over the weighting one (MAPA.W) to the relatively limited sample size.

It is also shown that MTA boosts forecasting performance, even when external variables that 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 sophisticated forecasting models.

Moreover, we demonstrate that our proposed approach achieves good accuracy even for limited sample sizes, being a fast, robust, and reliable solution. This is important given that, typically, forecasting must be performed automatically for numerous time series, to support decisions within an acceptable time frame. This conclusion is further supported by observing that the proposed approach can be easily implemented in energy management systems as well as existing forecasting support systems. It is based on exponential smoothing that is standard in most systems, offering a compelling alternative where more complex methods mentioned in the literature, such as machine learning techniques, are not available or applicable.

Finally, given the positive results reported here and in the literature for MTA, we suggest that future research should be focused on (i) optimally combining the forecasting methods across the multiple temporal aggregation levels, (ii) optimally combining temporal hierarchical levels to cross-sectional ones, and (iii) expanding cross-temporal aggregation for probabilistic forecasting.

## References

1. Jeanne, A.P., Mølgård, H.J.D., Kildegaard, D.N., Krogh, B.T.. Short-term balancing of supply and demand in an electricity system: forecasting and scheduling. *Annals of Operations Research* 2016;238(1):449–473.

2. Barzin, R., Chen, J.J., Young, B.R., Farid, M.M.. Peak load shifting with energy storage and price-based control system. *Energy* 2015;92, Part 3:505–514.

3. Biscarri, F., Monedero, I., Garca, A., Guerrero, J.I., Len, C.. Electricity clustering framework for automatic classification of customer loads. *Expert Systems with Applications* 2017;86:54–63.

4. Vu, D., Muttaqi, K., Agalgaonkar, A., Bouzerdoum, A.. Short-term electricity demand forecasting using autoregressive based time varying model incorporating representative data adjustment. *Applied Energy* 2017;205:790–801.

5. Adeoye, O., Spataru, C.. Modelling and forecasting hourly electricity demand in west african countries. *Applied Energy* 2019;242:311–333.

6. Tratar, L.F., Strmčnik, E.. The comparison of holtwinters method and multiple regression method: A case study. *Energy* 2016;109:266–276.

7. Amini, M.H., Kargarian, A., Karabasoglu, O.. Arima-based decoupled time series forecasting of electric vehicle charging demand for stochastic power system operation. *Electric Power Systems Research* 2016;140:378–390.

8. Ruiz, L., Rueda, R., Cullar, M., Pegalajar, M.. Energy consumption forecasting based on elman neural networks with evolutive optimization. *Expert Systems with Applications* 2018;92:380–389.

9. Cai, M., Pipattanasomporn, M., Rahman, S.. Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques. *Applied Energy* 2019;236:1078–1088.

10. Imani, M., Ghassemian, H.. Residential load forecasting using wavelet and collaborative representation transforms. *Applied Energy* 2019;253:113505.

11. Bedi, J., Toshniwal, D.. Deep learning framework to forecast electricity demand. *Applied Energy* 2019;238:1312–1326.

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

13. Ma, X., Jin, Y., Dong, Q.. A generalized dynamic fuzzy neural network based on singular spectrum analysis optimized by brain storm optimization for short-term wind speed forecasting. *Applied Soft Computing* 2017;54:296–312.

14. Yukseltan, E., Yucekaya, A., Bilge, A.H.. Forecasting electricity demand for turkey: Modeling periodic variations and demand segregation. *Applied Energy* 2017;193:287–296.

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

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

17. Bourdeau, M., qiang Zhai, X., Nefzaoui, E., Guo, X., Chatellier, P.. Modeling and forecasting building energy consumption: A review of data-driven techniques. *Sustainable Cities and Society* 2019;48:101533.

18. Kourentzes, N., Petropoulos, F., Trapero, J.R.. Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting* 2014;30(2):291–302.

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

20. Pedregal, D.J., Trapero, J.R.. Mid-term hourly electricity forecasting based on a multi-rate approach. *Energy Conversion and Management* 2010;51(1):105–111.

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

22. Zhang, Y., Dong, J.. Least squares-based optimal reconciliation method for hierarchical forecasts of wind power generation. *IEEE Transactions on Power Systems* 2018;:1–1.

23. Yang, D., Quan, H., Disfani, V.R., Liu, L.. Reconciling solar forecasts: Geographical hierarchy. *Solar Energy*

2017;146:276–286.

24. Hyndman, R.J., Ahmed, R.A., Athanasopoulos, G., Shang, H.L.. Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis* 2011;55(9):2579–2589.

25. Spiliotis, E., Petropoulos, F., Assimakopoulos, V.. Improving the forecasting performance of temporal hierarchies. *PLOS ONE* 2019;14(10):1–21.

26. Athanasopoulos, G., Ahmed, R.A., Hyndman, R.J.. Hierarchical forecasts for australian domestic tourism. *International Journal of Forecasting* 2009;25(1):146–166.

27. Gross, C.W., Sohl, J.E.. Disaggregation methods to expedite product line forecasting. *Journal of Forecasting* 1990;9(3):233–254.

28. Villegas, M.A., Pedregal, D.J.. Supply chain decision support systems based on a novel hierarchical forecasting approach. *Decision Support Systems* 2018;114:29–36.

29. Wickramasuriya, S.L., Athanasopoulos, G., Hyndman, R.J.. Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association* 2019;114(526):804–819.

30. Shlifer, E., Wolff, R.. Aggregation and proration in forecasting. *Management Science* 1979;25(6):594–603.

31. D'Attilio, D.F.. Practical applications of trend analysis in business forecasting. *The Journal of Business Forecasting Methods & Systems* 1989;8:9–11.

32. Dangerfield, B.J., Morris, J.S.. Top-down or bottom-up: Aggregate versus disaggregate extrapolations. *International Journal of Forecasting* 1992;8(2):233–241.

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

34. Schwarzkopf, A.B., Tersine, R.J., Morris, J.S.. Top-down versus bottom-up forecasting strategies. *International Journal of Production Research* 1988;26(11):1833–1843.

35. Zheng, Z., Chen, H., Luo, X.. A kalman filter-based bottom-up approach for household short-term load forecast. *Applied Energy* 2019;250:882–894.

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

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

38. Kohn, R.. When is an aggregate of a time series efficiently forecast by its past? *Journal of Econometrics* 1982;18(3):337–349.

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

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

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

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

43. Kavgic, M., Mavrogianni, A., Mumovic, D., Summerfield, A., Stevanovic, Z., Djurovic-Petrovic, M.. A review of bottom-up building stock models for energy consumption in the residential sector. *Building and Environment* 2010;45(7):1683–1697.

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

45. Lai, S.H., Hong, T.. When One Size No Longer Fits All: Electric Load Forecasting with a Geographic Hierarchy; 2013. SAS White Paper; URL http://assets.fiercemarkets.net/public/sites/energy/reports/electricloadforecasting.pdf.

46. Petropoulos, F., Hyndman, R.J., Bergmeir, C.. Exploring the sources of uncertainty: Why does bagging for time series forecasting work? *European Journal of Operational Research* 2018;268(2):545–554.

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

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

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

50. Petropoulos, F., Kourentzes, N.. Forecast combinations for intermittent demand. *Journal of the Operational Research Society* 2015;66(6):914924.

51. Petropoulos, F., Kourentzes, N.. Improving forecasting via multiple temporal aggregation. *Foresight: The International Journal of Applied Forecasting* 2014;2014(34):12–17.

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

53. Kourentzes, N., Petropoulos, F.. Forecasting with multivariate temporal aggregation: The case of promotional modelling. *International Journal of Production Economics* 2016;181, Part A:145–153.

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

55. Kourentzes, N., Rostami-Tabar, B., Barrow, D.K.. Demand forecasting by temporal aggregation: using optimal or multiple aggregation levels? *Journal of Business Research* 2017;78:1–9.

56. Yang, D., Quan, H., Disfani, V.R., Rodrguez-Gallegos, C.D.. Reconciling solar forecasts: Temporal hierarchy. *Solar Energy* 2017;158:332–346.

57. Kourentzes, N., Athanasopoulos, G.. Cross-temporal coherent forecasts for australian tourism. *Annals of Tourism Research* 2019;75:393–409.

58. Yagli, G.M., Yang, D., Srinivasan, D.. Reconciling solar forecasts: Sequential reconciliation. *Solar Energy*

2019;179:391–397.

59. Martnez, F., Fras, M.P., Prez-Godoy, M.D., Rivera, A.J.. Dealing with seasonality by narrowing the training set in time series forecasting with knn. *Expert Systems with Applications* 2018;103:38–48.

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

61. Miller, D.M., Williams, D.. Shrinkage estimators of time series seasonal factors and their effect on forecasting accuracy. *International Journal of Forecasting* 2003;19(4):669–684.

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

63. Ledolter, J.. The effect of additive outliers on the forecasts from arima models. *International Journal of Forecasting* 1989;5(2):231–240.

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

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

66. Kendall, M., Stuart, A.. The advanced theory of statistics. *Griffin* 1983;3:410–414.

67. Beaumont, A.N.. Data transforms with exponential smoothing methods of forecasting. *International Journal of Forecasting* 2014;30(4):918–927.

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

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

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

71. Hyndman, R.J., Koehler, A.B.. Another look at measures of forecast accuracy. *International Journal of Forecasting* 2006;22(4):679–688.

72. Tashman, L.J.. Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting* 2000;16(4):437–450.

73. R Core Team, . R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria; 2018. URL https://www.R-project.org/.

74. Kourentzes, N., Petropoulos, F.. MAPA: Multiple Aggregation Prediction Algorithm; 2018. R package version 2.0.4; URL https://CRAN.R-project.org/package=MAPA.

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

76. de Lacalle, J.L.. tsoutliers: Detection of Outliers in Time Series; 2017. R package version 0.6-6; URL https://CRAN.R-project.org/package=tsoutliers.

77. Crdenas, J.J., Romeral, L., Garcia, A., Andrade, F.. Load forecasting framework of electricity consumptions for an intelligent energy management system in the user-side. *Expert Systems with Applications* 2012;39(5):5557–

5565.

78. Martinez-Pabon, M., Eveleigh, T., Tanju, B.. Optimizing residential energy management using an autonomous scheduler system. *Expert Systems with Applications* 2018;96:373–387.

79. Pea, M., Biscarri, F., Guerrero, J.I., Monedero, I., Len, C.. Rule-based system to detect energy efficiency anomalies in smart buildings, a data mining approach. *Expert Systems with Applications* 2016;56:242–255.

80. Spiliotis, E., Legaki, N.Z., Assimakopoulos, V., Doukas, H., El Moursi, M.S.. Tracking the performance of photovoltaic systems: a tool for minimising the risk of malfunctions and deterioration. *IET Renewable Power Generation* 2018;12(7):815–822.

81. Makridakis, S., Hibon, M.. The M3-Competition: results, conclusions and implications. *International Journal of Forecasting* 2000;16(4):451–476.

82. Spiliotis, E., Kouloumos, A., Assimakopoulos, V., Makridakis, S.. Are forecasting competitions data representative of the reality? *International Journal of Forecasting* 2020;36(1):37–53.