Department of Econometrics
Erasmus School of Economics
Erasmus University Rotterdam

Bachelor Thesis
BSc Econometrics & Operations Research

July 2, 2017

USING THE MULTIPLICATIVE DOUBLE SEASONAL HOLT-WINTERS METHOD TO FORECAST SHORT-TERM ELECTRICITY DEMAND

by:
Aagje Franse (414567)

Thesis Supervisor: Dr. Mikhail Zhelonkin
Co-reader: Dr. Michel van de Velden

Abstract

Balancing electricity demand and supply is a challenging task for the National Grid. Accurate short-term electricity demand forecasts are of vital importance to ensure that electricity can be supplied according to demand without wasting financial and energy resources. Time series data on electricity demand in England and Wales, collected at half-hourly intervals, reveals both a within-day and a within-week seasonal cycle. In this paper, two new univariate short-term forecasting methods accommodating two seasonal cycles are proposed to model electricity demand: double seasonal Holt-Winters exponential smoothing and singular spectrum analysis. The methods are compared to a benchmark seasonal ARIMA model in terms of forecasting performance as measured by the mean absolute percentage error. The double seasonal Holt-Winters method with a simple autoregressive model fitted to its residuals outperforms the benchmark model; its mean absolute percentage errors are significantly smaller than those of the seasonal ARIMA model. In contrast, the method of singular spectrum analysis outperforms neither the benchmark nor the seasonal ARIMA model.

Contents

1 Introduction

The National Grid operates the transmission network that connects power stations to electricity consumers in England and Wales. The managing and controlling of this power system is a challenging task, as energy cannot be stored, and hence, has to be generated according to demand in order to avoid the wastage of energy sources.

Bilateral trading between generator sites and suppliers enables electricity supply companies to meet their customer's demand. The National Grid is responsible for balancing the demand and supply system and ensuring that power stations will be able to provide electricity in case of an unanticipated increase in electricity demand after the closure of bilateral contracts. For this, the National Grid goes to the so-called 'balancing market', which operates on a rolling basis 1 hour ahead of real-time, where generating stations will be ready to cater to the unmet demand for electricity[1] (Taylor, 2003). Therein, a challenge lies: on the one hand, having too many generating sites on stand-by means a substantial loss of both financial and energy resources. On the other hand, too little generating sites in readiness could cause the frequency to drop below the nominal system frequency of 50.00 Hz (Bunn, 1982). To avoid either an over- or an undersupply of electricity, short-term electricity demand forecasts are of vital importance to ensure that supply and demand can be balanced real time. At the National Grid, an online prediction mechanism based on half-hourly data is used to forecast load for lead times ranging up to a day-ahead (Taylor, 2003).

Multivariate methods are uncommon when forecasting short-term electricity demand, due to their large input requirements and lack of robustness (Bunn, 1982). Hence, only univariate forecasting methods are considered for the short-term load forecasting in this paper.

The data on half-hourly demand for electricity in England and Wales reveals two seasonal patterns; one within-day seasonal pattern consisting of 48 subsequent half-hour periods, and one within-week seasonal pattern consisting of 336 subsequent half-hour periods (Taylor, 2003).

The multiplicative ARIMA model has previously been extended to accommodate for more than one seasonal cycle. However, in this paper, we aim to develop a new online forecasting method to capture both the within-day and within-week seasonal cycle in electricity demand forecasting. A method frequently used when forecasting seasonal time series is Holt-Winters exponential smoothing. In this research, we will extend standard Holt-Winters exponential smoothing to accommodate for both the within-day and the within-week seasonal cycle. Afterwards, the performance of this new double seasonal Holt-Winters method is evaluated. The aim of this paper is to improve the accuracy of the forecasts made using the double seasonal Holt-Winters method, in comparison to the forecasting accuracy of the standard Holt-Winters method. We hypothesise that the double seasonal Holt-Winters method will outperform the standard Holt-Winters method as well as a benchmark multiplicative ARIMA model. Lastly, the method of singular spectrum analysis is proposed to model time series containing two seasonal cycles. This method is reviewed in terms of forecasting performance in comparison to the standard Holt-Winters method, the double seasonal Holt-Winters methods and the benchmark model. We hypothesise that the singular spectrum analysis will outperform all of the aforementioned methods.

The outline of the paper is as follows. In Section 2 the related literature is discussed. A short discussion of the data is presented in Section 3. Afterwards, the benchmark model and double seasonal Holt-Winters model are introduced in Sections 4.1 and 4.2 and an alternative method of time series analysis and forecasting is presented in Section 4.3. The results and comparison of the forecasting performance are presented in Section 5 followed by a discussion with suggestions for further research in Section 6 and a conclusion in Section 7.

[1]Information on the National Grid was extracted from its official website: http://www2.nationalgrid.com/uk/

2 Literature Review

In this section the relevant literature regarding our study is reviewed. The section is organised as follows: first, we will briefly touch upon the history of time series prediction in Section 2.1, afterwards there will be a closer examination of load forecasting in Section 2.2 and lastly the relevant literature of the methods applied in this research will be discussed.

2.1 Time Series Forecasting

The analysis of time series can be traced back to 1919, when Warren Persons distinguished between four time series components: the trend, the business cycle, the seasonal pattern and the residual (Kirchgässner and Wolters, 2007). The classic approach to time series modelling is very closely related to regression analysis, due to the fact that it was not a common idea at the time that observations might be chronologically dependent upon one another (Kirchgässner and Wolters, 2007). One of the first time series-related findings dates back to 1949, when Cochrane and Orcutt (1949) found that the error terms of a linear regression model could be autocorrelated, causing the variance of the estimator and the t- and F-statistics to be affected. Simultaneously, the Durbin-Watson test statistic was developed to help detect autocorrelation in the residuals of a regression (Durbin and Watson, 1950). Later, Box and Jenkins (1970) developed a three-stage approach to build univariate models that could forecast time series data only using past observations. This modern approach to time series forecasting gained popularity quickly, partly due to the work of Granger and Newbold (1975), who proposed that forecasts made using past time series observation usually outperformed forecasts made by larger and more complex econometric models (Kirchgässner and Wolters, 2007, and references therein). The Box-Jenkins modelling approach is common as a benchmark model up to this point.

2.2 Load Forecasting

Load forecasting can be divided into very short-term, short-term, medium-term and long-term load forecasting, depending on the time horizon of the forecast. Very short-term load forecasting (VSTLF) predicts load from one-minute ahead up to one-hour ahead. The forecast horizon of short-term load forecasting (STLF) is longer, covering periods that last up to one week, while the period for medium-term load forecasting (MTLF) usually ranges between one week and one year. Lastly, long-term load forecasting (LTLF) is used for periods longer than one year (Badran et al., 2008). The focus of this paper is STLF, which is most applicable to concerns such as balancing the market after the closure of bilateral contracts and daily planning and scheduling of the generator sites.

Extensive research has been carried out on the topic of univariate short-term load forecasting, covering a broad spectrum of different forecasting techniques (Peng and Ching-Wu, 2009; Bunn, 1982, and references therein). An elaborate overview of different univariate forecasting methods can be found in Bunn (1982). In this paper Bunn presents a review on several important aspects of various short-term forecasting techniques for the electricity industry, such as adaptiveness, recursiveness, the computational economy and robustness. It states that there is no optimal method to forecast short-term electricity demand and that choice of method depends on its logic and implementability. Finally, we will take a closer look at the available methods proposed in important works and their related findings. All three methods discussed below are applied in this paper.

ARIMA Modelling

The three-step modelling approach proposed by Box and Jenkins (1970) consists of a model identification, estimation and model checking stage. This approach has been highly popular ever since

1970, due to the broad applicability of the method and the large possibility to find an adequate ARIMA model for the data (Hyndman, 2001). Besides, Box et al. (1967) developed the so-called “airline model”, which was later generalized to a class of multiplicative ARIMA models containing a seasonal pattern: the seasonal ARIMA (SARIMA) models (Box and Jenkins, 1970). Nowadays, it is not uncommon for the multiplicative SARIMA model to be used as a benchmark approach, see for example Abraham and Baikunth (2001) or Abu-El-Magd and Sinha (1982). The SARIMA model will serve as a benchmark model in our research, and is discussed in further detail in Section 4.1.

Exponential Smoothing
Exponential smoothing is a common approach to short-term univariate load forecasting due to its accuracy, relatively easy implementability and simplicity (Christiaanse, 1971). Holt (2004) extended simple exponential smoothing to enable forecasting in case the data contains a trend (Hyndman and Athanasopoulos, 2001). Later Winters (1960) extended the method further to enable forecasting of time series containing a seasonal cycle, now known as Holt-Winters exponential smoothing. In our research Holt-Winters exponential smoothing will be extended to capture two seasonal patterns, as opposed to merely one. Over the past decades several new exponential smoothing techniques have been proposed. Taylor (2012) reviews the performance of a new singular value decomposition-based (SVD-based) exponential smoothing technique in comparison to several other weighted exponential smoothing methods; however, it is not found to outperform the more traditional Intraday Cycle (IC) exponential smoothing.

Singular Spectrum Analysis
The non-parametric method of singular spectrum analysis (SSA) has gained more popularity over the last few decades. The method is widely applicable and has been used for a wide range of subjects, including quantitative genetics studies (Hassani and Ghodsi, 2015), short-term traffic flow prediction (Shang et al., 2016) and image processing (Rodrguez-Aragn and Zhigljavsky, 2010), but can be applied to economics, finance and social sciences as well. A complete description of the theory behind SSA and related techniques is presented in Golyandina et al. (2001) and in Golyandina and Zhigljavsky (2013). Schoellhamer (2001) investigates the analysis of geographic time series with missing data using SSA and finds that SSA can be applied successfully to time series with missing data, whereas the more recent study of Alexandrov (2009) presents a method of trend extraction for time series using SSA.

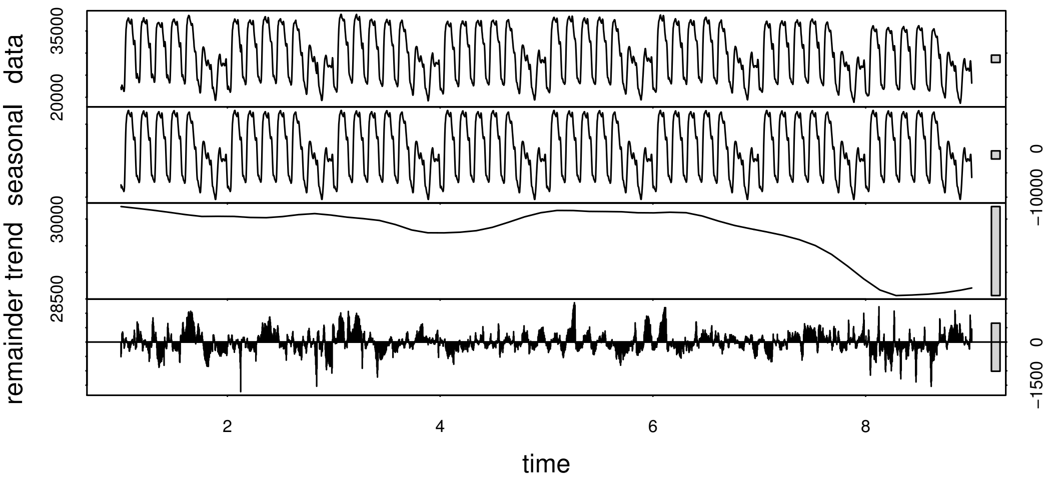
3 Data

The data available for our research consists of 12 weeks of half-hourly electricity demand in Wales and England and was collected during the 84 days between 5 June 2000 and 27 August 2000. These 4032 data points are split up into two samples: one sample containing 2688 data points collected in the first 8 weeks, hereafter called the ‘training set’, and another sample containing the remaining 1344 data points collected in the last 4 weeks, hereafter called the ‘testing set’. The training set is used to fit the model and estimate model parameters, whereas the second sample is used to evaluate the model’s forecasting performance and accuracy.

An seasonal and trend decomposition using Loess (STL) decomposition of the training set can be found in Figure 1, where the values on the x-axis represent the 8 weeks of data that make up the training set. As can be seen in the STL decomposition, the largest part of the variation in the data is caused by the seasonal components of the data. The range bars on the right side of the figure indicate how much of the variation in the data can be attributed to each component, where a relatively large bar as in comparison to the “data” bar indicates that less variation can be attributed to this component. As a result it can be concluded that the largest part of the variation in data is caused by the seasonal cycles, a smaller part of the variation is caused by the remainder and the smallest part of the variation can be attributed to the trend component.

Our research is simplified due to the fact that there are no so-called 'special days' occurring in the period spanned by our data; days in which the demand is so unlike the rest of the year that forecasting the demand for these days is nearly impossible and thus, these days require individual treatment (Taylor, 2003). Examples of such days are public and consecutive holidays such as Christmas or Liberation Day. When special days are present in the data, rule-based procedures that override the system off-line can be used to estimate the electricity demand for those days (Hyde and Hodnett, 1993).

Figure 1: STL decomposition of the training set (existing of the first 2688 observations of the data), where the time on the x-axis is in weeks.



4 Methodology

Here, the methodology is presented. The organisation of the section is as follows: first, the benchmark seasonal ARIMA model is presented in Section 4.1. Afterwards, the standard and double seasonal Holt-Winters methods are explained formally in Section 4.2, alongside a description of the method of parameter estimation. The method of singular spectrum analysis is proposed in Section 4.3, as well as a motivation for the choice of using it. Lastly, the method of comparing the forecasting accuracy of the methods is introduced in Section 4.4.

4.1 Benchmark Model

The methodology of the benchmark multiplicative seasonal ARIMA model and the double seasonal Holt-Winters method are presented in this section. Furthermore the parameter estimation and the implementation of the methods will be discussed.

4.1.1 Multiplicative Seasonal ARIMA

The multiplicative seasonal ARIMA model can be expressed as ARIMA$(p,d,q) \times (P,D,Q)_m$, where m is the number of periods in a seasonal cycle, p and P are the orders of the autoregressive parts, d and D are the degrees of differencing, and q and Q are the orders of the moving average parts of the model. The backshift notation of this ARIMA model capturing one seasonal cycle is as follows:

$$\phi_p(L)\Phi_P(L^m) \bigtriangledown^d \bigtriangledown_m^D y_t = \theta_q(L)\Theta_Q(L^m)e_t, \tag{1}$$

for a time series y_t with error term e_t, where L is the lag operator, ϕ_p, Φ_P, θ_q and Θ_Q are the lag polynomials. $\bigtriangledown$ and $\bigtriangledown_m$ are the normal and the seasonal difference operators: $(1 - L)$ for one lag, and $(1 - L^m)$ for the seasonal lag. In case of non-stationarity, the difference operators can transform the time series into a stationary time series by taking differences. To do so, the normal difference operator, $(1 - L)$ takes the difference between consecutive observations y_t and y_{t-1}, whereas the seasonal difference operator takes the difference between an observation and its corresponding observation from the previous season: y_t and y_{t-m}. In (1), the non-seasonal components of the model are given by $\phi_p(L)$, representing the non-seasonal AR terms and by $\theta_q(L)$, representing the non-seasonal MA terms. The seasonal components of the model are given by $\Phi_P(L^m)$ and $\Theta_Q(L^m)$ in (1), where the former represents the seasonal AR terms and the latter the seasonal MA terms. To determine the order of the seasonal ARIMA model we make use of the Bayesian Information Criterion (BIC), comparing models based on both the original training data and the differenced data. As the seasonal cycle of length $m = 336$ constitutes the largest part of the seasonal component, the seasonality of the SARIMA model is set to 336. We consider AR and MA lags up to order 4, and calculate the BIC for all possible SARIMA models within this range. All calculated BICs are shown in Table 3 (undifferenced data) and in Table 4 (differenced data) in Appendix A. After determining the order of the benchmark model, the parameters are estimated using the maximum likelihood (ML) method. The model with the lowest BIC was the following ARIMA$(2, 1, 2) \times (0, 1, 0)_{336}$ model, hereafter called the 'seasonal ARIMA' model:

$$(1 + 0.34L^2)(1 - L) \times (1 - L^{336})(Y_t - 0.008) = (1 + 0.38L^2)\epsilon_t \tag{2}$$

A possible explanation for the fact that only the second AR and MA lags are incorporated in this model is the fact that although demand is measured by the National Grid every half-hour, the demand for electricity has an hourly dependence as opposed to a half-hourly dependence.

4.2 Holt-Winters Exponential Smoothing

Depending on the type of seasonal pattern present in the data one can opt for two different varieties of Holt-Winters exponential smoothing: namely, multiplicative and additive exponential smoothing (Hyndman and Athanasopoulos, 2001). In the additive model, the magnitude of the seasonal component does not depend on the overall level of the series, and a certain value can be added to the series to account for seasonality. Hence, the nature of this seasonality is additive. Alternatively, in the multiplicative model the magnitude of the seasonal component is relative to the magnitude of the series, and hence, the level of the series can be multiplied by a certain factor to account for this seasonality. This implies that the nature of the seasonality is multiplicative (Kalekar, 2004).

One could derive the nature of the seasonality present in one's data by inspecting the plotted data. However, as the time span of the data available to us is a mere 12 weeks consists of only 3042 observations, distinguishing between multiplicative and additive seasonality becomes rather complex. In our research, we ultimately opt for the multiplicative Holt-Winters method for the following two reasons. Firstly, the multiplicative model is more widely used than the additive model and thus an improvement of this method will be more valuable than an improvement of the additive model Taylor (2003). Furthermore, Taylor (2003) investigates the multiplicative Holt-Winters exponential smoothing method in his research, as opposed to the additive Holt-Winters exponential smoothing method. Since we aim to reproduce Taylor's results, it is suitable to use an approach that is similar to his approach.

4.2.1 Standard Holt-Winters Exponential Smoothing

The standard multiplicative Holt-Winters exponential smoothing method is capable of accommodating one seasonal pattern. It comprises of one forecast equation for a k-step ahead forecast, as well as three smoothing equations for respectively the level, trend and seasonality of the series:

$$\hat{Y}_t(k) = (S_t + kT_t)I_{t-m+k}, \tag{3}$$

$$S_t = \alpha(Y_t/I_{t-m}) + (1-\alpha)(S_{t-1} + T_{t-1}), \tag{4}$$

$$T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1}, \tag{5}$$

$$I_t = \delta(Y_t/S_t) + (1-\delta)I_{t-m}, \tag{6}$$

where α, γ and δ are the smoothing parameters. Furthermore, the assumption of an additive trend is made in equations (3) - (6). In (4), the level index at time t, S_t, is estimated by smoothing the ratio of the current observation, Y_t, to the seasonality index of the corresponding observation of the previous season, I_{t-m}. Likewise, in (5), the trend index, T_t is estimated by smoothing the level differences, $(S_t - S_{t-1})$, whereas in (6), the seasonality index, I_t, is estimated by smoothing the ratio of the current observation, Y_t, to the level index, S_t.

We produce forecasts using two varieties of the standard multiplicative Holt-Winters methods: standard Holt-Winters capturing for the within-day seasonal cycle and standard Holt-Winters for the within-week seasonal cycle. For the former, the value of m is set to 48, whereas for the latter, the value of m is set to 336. The set of initial values for both methods consists of a scalar value for the initial trend and level index, and a column vector of length m for the initial seasonality indices. In consistence with the method proposed by Hyndman (1998), the initial values are determined as follows:

$$S_0 = \frac{(y_1 + y_2 + \cdots + y_m)}{m}, \tag{7}$$

$$T_0 = \frac{(y_{m+1} + y_{m+2} + \cdots + y_{m+m}) - (y_1 + y_2 + \cdots + y_m)}{m^2}, \tag{8}$$

$$I_0 = \begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_m \end{bmatrix} = \begin{bmatrix} y_1/S_0 \\ y_2/S_0 \\ \vdots \\ y_m/S_0 \end{bmatrix}. \tag{9}$$

After setting the initial values, forecasts can be constructed using (3) - (6). The smoothing parameters α, γ and δ are estimated by minimising the sum of squared errors of prediction (SSE) of the one-step ahead forecasts of the observations in the training set, using the MATLAB *fmincon* function. The expression used to calculate the SSE is as follows:

$$SSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \tag{10}$$

where N is the number of observations for which one-step ahead forecasts are made. Estimated parameters are shown in Tables 1 and 2 in Section 5.1.

4.2.2 Double Seasonal Holt-Winters Exponential Smoothing

Since the standard Holt-Winters exponential smoothing method merely captures one seasonal cycle, we have to extend the model to enable it to capture two seasonal cycles. We do this by applying the method proposed by Taylor (2003): by including two separate equations for both the within-day and the within-week seasonality index. The Holt-Winters method could be extended to accommodate three or more seasonal cycles in a similar fashion. The double seasonal multiplicative Holt-Winters exponential smoothing method is given by the following equations:

$$\hat{Y}_t(k) = (S_t + kT_t)D_{t-m_1+k}W_{t-m_2+k}, \tag{11}$$

$$S_t = \alpha\{Y_t/(D_{t-m_1}W_{t-m_2})\} + (1-\alpha)(S_{t-1} + T_{t-1}), \tag{12}$$

$$T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1}, \tag{13}$$

$$D_t = \delta\{Y_t/(S_tW_{t-m_2})\} + (1-\delta)D_{t-m_1}, \tag{14}$$

$$W_t = \omega\{Y_t/(S_tD_{t-m_1})\} + (1-\omega)W_{t-m_2}, \tag{15}$$

for a k-step ahead forecast. α, γ, δ and ω are the smoothing parameters, whereas D_t and W_t are seasonal indices for seasonal cycles of length m_1 and m_2, respectively. At time t, the level index, S_t, is calculated by smoothing the current observation, Y_t, to the product of the within-day and within-week seasonality indices of the corresponding observations in the previous seasonal cycles, $D_{t-m_1}W_{t-m_2}$ (12). Furthermore, the trend, T_t, is determined smoothing the successive differences, $(S_t - S_{t-1})$, of the level (13). By smoothing the current observation, Y_t, to the product of the level index and the within-week seasonal index of the corresponding observation in the previous season, $S_tW_{t-m_2}$, the within-day seasonal index, D_t, is determined (14). Lastly, the within-week seasonality, W_t, is calculated by smoothing the current observation, Y_t, to the level index and the within-day seasonality index of the corresponding observation in the previous season, $S_tD_{t-m_1}$ (15).

For the double seasonal Holt-Winters method, the initial values, S_0, T_0, D_0 and W_0 are specified as described in Taylor (2003).

After setting the initial values, forecasts can be constructed. The smoothing parameters α, γ, δ and ω are estimated by minimising the SSE of the one-step ahead forecasts of the values in the training set, where the SSE is defined as in (10). The estimated parameters are presented in Section 5.1 in Tables 1 and 2, alongside with an interpretation of the coefficients.

4.2.3 First-Order Autocorrelation Adjustment

After careful inspection of the sample autocorrelation plots of within-day, within-week and double seasonal Holt-Winters exponential smoothing, a significant first-order autocorrelation is detected for all three methods (see also Figures 11, 12 and 13 in Appendix B for the sample autocorrelation plots). Gardner (1985) states that predictions can be modified by a small adjustment for the autocorrelation in this case, as illustrated by Chatfield (1978) and Reid (1975). To adjust the Holt-Winters methods for first-order autocorrelation, we add an extra term to the expressions (3) and (11). This method was first proposed in Gilchrist (1976), and involves fitting an AR(1) model to the one-step ahead forecast errors of the model. The forecasting expressions then become:

$$\hat{Y}_t(k) = (S_t + kT_t)I_{t-m+k} + \phi^k\{Y_t - (S_{t-1} + T_{t-1})I_{t-m}\}, \tag{16}$$

for Holt-Winters exponential smoothing accommodating one seasonal cycle, and:

$$\hat{Y}_t(k) = (S_t + kT_t)D_{t-m_1+k}W_{t-m_2+k} + \phi^k\{Y_t - (S_{t-1} + T_{t-1})D_{t-m_1}W_{t-m_2}\}, \tag{17}$$

for double seasonal Holt-Winters exponential smoothing.

4.3 Singular Spectrum Analysis

Singular Spectrum Analysis (SSA) is a non-parametric method that can be used to analyse and forecast time series (Golyandina et al., 2001) that satisfy a Linear Recurrent Formula (LRF). The underlying idea of singular spectrum analysis is that a time series can be decomposed into several time series components, such as trend, seasonality and the like, using singular value decomposition (SVD). These components can then be used to generate 'new' time series and to use these to forecast the time series. SSA has already proven itself to be a powerful and widely applicable method in several different fields of research, see also Section 2.2. One of the main benefits of SSA is that the time series can be decomposed into several independent and interpretable components, making analysis of the time series less complex. Another benefit is that SSA can be applied to complex time series for a variety of applications; examples of those are trend extraction, smoothing and the extraction of different seasonal cycles (these, and more, applications are demonstrated in Hassani (2007)). Furthermore, in SSA no assumptions regarding the signal and noise of the time series are made (Golyandina et al., 2001). Due to its promising results in the past, wide applicability and ease of interpretation SSA will be implemented in order to investigate whether it can compete with, or even outperform the more traditional Holt-Winters exponential smoothing methods.

The subsection is divided as follows: in Section 4.3.1, the theory of SSA will be discussed briefly, following the approach of Golyandina et al. (2001). Afterwards the forecasting method will be examined in Section 4.3.2 and the implementation of SSA in Sections 4.3.3 and 4.3.4.

4.3.1 Brief review of SSA

Decomposition: Embedding
In the first step of SSA, a time series $y_1, y_2, \cdots, y_N$ is mapped onto a series of lagged vectors to construct a so-called *trajectory matrix.* Let L be an integer, where $1 < L < N$, and $K = N - L + 1$. Then, the L-lagged vectors can be constructed as follows:

$$Y_i = \begin{bmatrix} y_i \\ y_{i+1} \\ \vdots \\ y_{i+L-1} \end{bmatrix}, \tag{18}$$

and the trajectory matrix of the series is given by:

$$\mathbf{Y} = \begin{bmatrix} Y_1 & Y_2 & \cdots & Y_K \end{bmatrix} = \begin{bmatrix} y_1 & \cdots & y_K \\ \vdots & \ddots & \vdots \\ y_L & \cdots & y_N \end{bmatrix}. \tag{19}$$

Since the elements on the diagonals of the matrix are equal, the trajectory matrix $\mathbf{Y}$ is a *Hankel Matrix.* The choice of parameter L will be discussed further in Section 4.3.3, however, a typical choice for L is an integer that is relatively close to $\frac{N}{2}$ and, if seasonal cycles are present in the time series, a multiple of the length of these seasonal cycles.

Decomposition: Singular Value Decomposition
In this step, the trajectory matrix is decomposed using SVD. To do so, the matrix $\mathbf{YY}^T$ is determined. Afterwards, the eigentriples (λ_i, U_i, V_i) of $\mathbf{YY}^T$ are determined, where λ_i denotes the eigenvalues $\lambda_1 \leq \cdots \leq \lambda_L \leq 0$ of $\mathbf{YY}^T$ in decreasing order and U_i denotes the corresponding system of eigenvectors $U_1, \cdots, U_L$ $(i = 1, 2, \cdots, L)$. Now, if we set $d = \text{rank}(\mathbf{Y})$, we can set the i-th factor vector $V_i = \mathbf{Y}^T U_i / \sqrt{\lambda_i}$ $(i = 1, 2, \cdots, d)$ and rewrite the elements $\mathbf{Y}_i$ of the SVD as:

$$\mathbf{Y}_i = \sqrt{\lambda_i} U_i V_i^T. \tag{20}$$

Finally, the SVD of the trajectory matrix can be written as:

$$\mathbf{Y} = \mathbf{Y}_1 + \mathbf{Y}_2 + \cdots + \mathbf{Y}_d. \tag{21}$$

Reconstruction: Eigentriple Grouping
Once the decomposition in (21) has been made, a grouping of the eigentriples is constructed. The aim of this grouping is to distinguish between the several additive components of the time series, according to their separability. The process of eigentriple grouping for our research are discussed in further detail in Section 4.3.3, while the theoretic basics are discussed in this section. During the process of eigentriple grouping, the set of indices $1, 2, \cdots, d$ is split up in M subsets $I_1, I_2, \cdots, I_M$. Correspondingly, the elementary matrices $\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_d$ are split into M subgroups as well, resulting in the following decomposition:

$$\mathbf{Y} = \mathbf{Y}_{I_1} + \mathbf{Y}_{I_2} + \cdots + \mathbf{Y}_{I_M}. \tag{22}$$

Reconstruction: Diagonal Averaging
In the last part of the reconstruction, each matrix $\mathbf{Y}_{I_i}$, for $i = 1, 2, \cdots, M$, is transformed into a new times series of length N. To do so, we set $\mathbf{L}$: an $L \times K$ matrix with elements y_{ij}, where $y^*_{ij} = y_{ij}$ if $L < K$ and $y^*_{ij} = y_{ji}$ otherwise. Furthermore, we set $L^* = \min(L,K)$ and $K^* = \max(L,K)$. The time series $g_0, g_1, \cdots, g_{N-1}$ can then be constructed using *diagonal averaging*:

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} y^*_{m,k-m+2}, & \text{for } 0 \leq k < L^* - 1, \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y^*_{m,k-m+2}, & \text{for } L^* - 1 \leq k < K^*, \\ \frac{1}{N-K} \sum_{m=k-K^*+2}^{N-k^*+1} y^*_{m,k-m+2}, & \text{for } K^* \leq k < N. \end{cases} \tag{23}$$

4.3.2 Recurrent Forecasting Algorithm Based on Basic SSA

We will now provide a global and intuitive description of the recurrent forecasting algorithm used to produce forecasts. A more extensive formal description of the algorithm can be found in Golyandina et al. (2001).

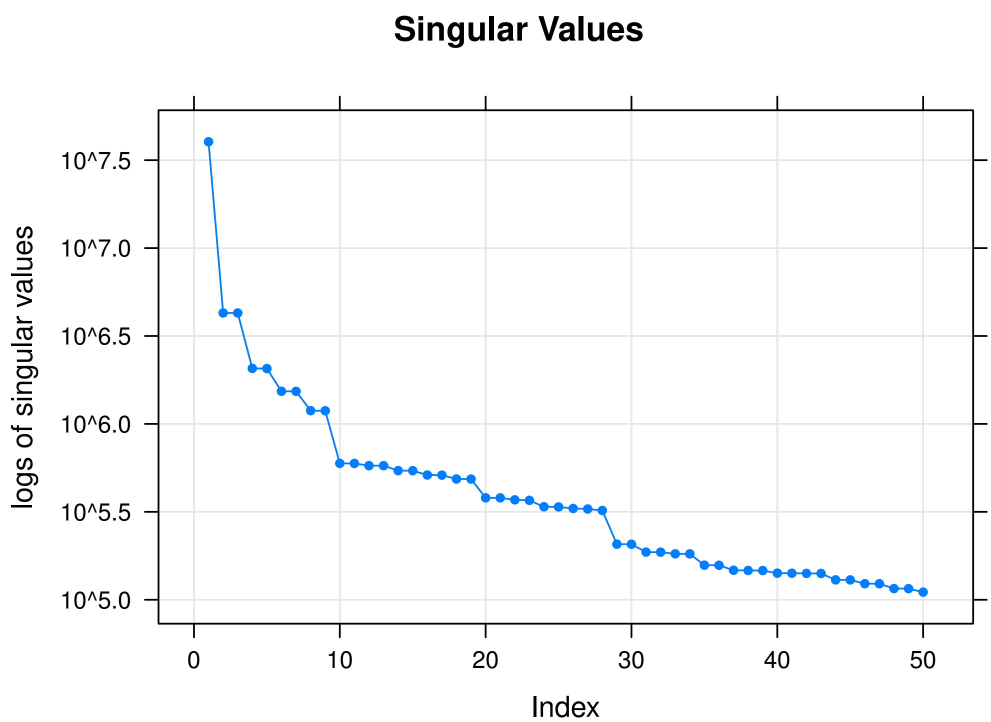
After applying basic SSA to our training set, we are left with $K = (N - L + 1)$ L-lagged vectors $Y_1, Y_2, \cdots, Y_K$ for any group of indices I_j selected in the eigentriple grouping stage of the reconstruction ($j = 1, 2, \cdots, M$). These vectors lie in an r-dimensional subspace of $\mathbb{R}^{\mathrm{L}}$, where r is the amount of elements in I_j. Furthermore, for $i = 1, 2, \cdots, K$, $\hat{Y}_i$ is the projection of the L-lagged vector Y_i onto the subspace $\mathcal{L}_r$. This subspace is spanned by r eigenvectors of the trajectory matrix $\mathbf{Y}$, previously discussed in Section 4.3.1. In the recurrent forecasting method, the vectors $\hat{Y}_1, \hat{Y}_2, \cdots, \hat{Y}_K$ are continued H steps with the continuation vectors $Z_{K+1}, Z_{K+2}, \cdots, Z_{K+H}$. The continuation vectors belong to subspace $\mathcal{L}_r$ and the resulting forecasting matrix $[\hat{Y_1}; \hat{Y_2}; \cdots; \hat{Y_K}; Z_{K+1}; Z_{K+2}; \cdots, Z_{K+H}]$ should approximately be a Hankel matrix. To complete the recurrent forecasting procedure, we perform diagonal averaging on the forecasting matrix to obtain the forecasted series (Golyandina et al., 2001).

4.3.3 Implementation of SSA

We use the 'RSSA' package for R by Korobeynikov, Golyandina, Usevich and Shlemov to implement SSA and the recurrent forecasting method. The first important step in performing basic SSA on the training set is the choice of parameter L. As mentioned above, L should be as large as possible, but no larger then $\frac{N}{2}$. If there are seasonal cycles present, L should be a multiple of those seasonal cycles. In order to satisfy the conditions for L, $\frac{N}{2}$ is set as the value for L, where N is the number of available observations of the time series. Now, the decomposition step of SSA can be performed. After decomposing the time series, the eigentriples have to be grouped according to which component of the time series they represent, in order to be able to perform the reconstruction step. For this grouping procedure, the **w**-correlation plot, the singular values and the phase plot are used.

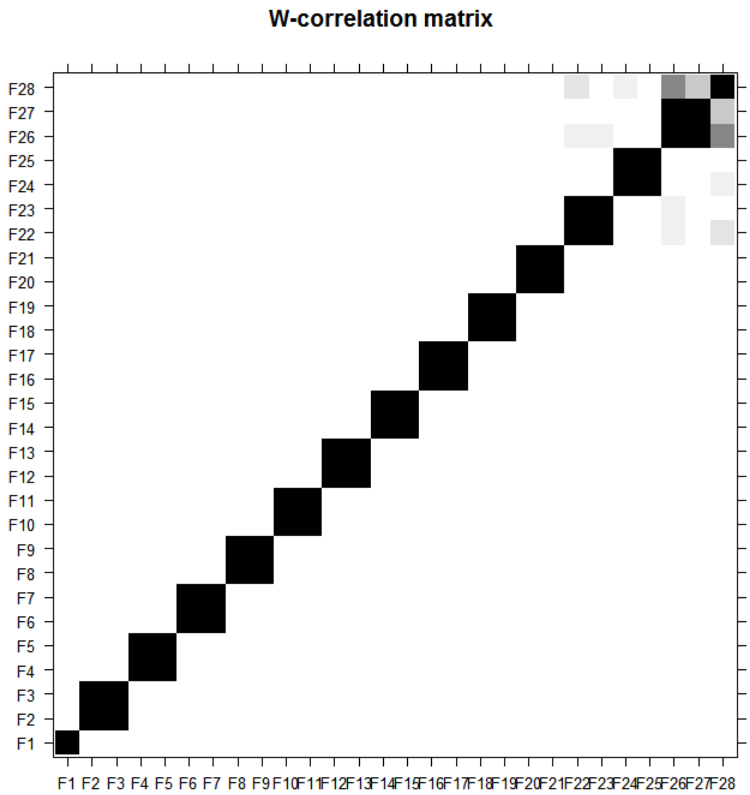
Singular values plot: A plot of the singular values helps us to identify groups of eigentriples representing the independent components of the time series. In Figure 2, the singular values plot of the training set after SVD is shown. This figure indicates the pairing of the eigentriples 2 and 3; 4 and 5; 6 and 7 and 8 and 9.

Figure 2: Scree plot of singular values after the decomposition stage of SSA.



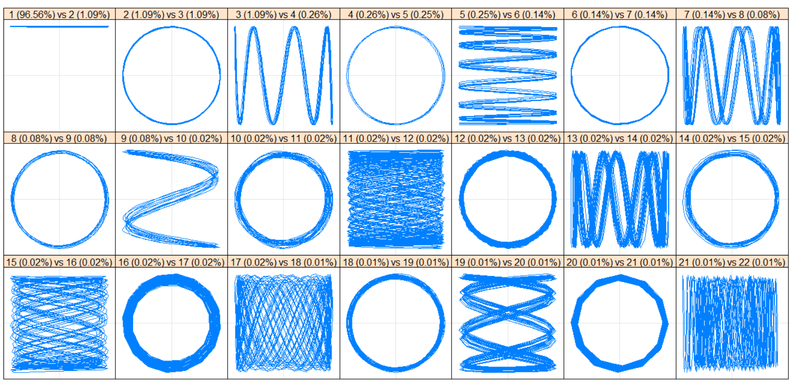
W-correlation matrix: The **w**-correlation matrix gives a good indication of the separability of the different eigentriples, and thus of the components of the time series. Figure 3 indicates a separation between, on one hand, the trend and harmonics components of the time series and, on the other hand, the white noise components. We can thus split the components into two groups: one group representing a smoothed version of the original time series, containing the first 21 eigentriples, and one group representing the residuals, containing the remaining eigentriples.

Figure 3: **W**-correlation plot of the first 28 eigentriples of the decomposition step of SSA



Phase plot: Phase plots of pairs of components are helpful when identifying the harmonic components of the time series. In Figure 4, pairs 2 and 3; 4 and 5; 6 and 7; 8 and 9; 10 and 11; 12 and 13; 14 and 15; 16 and 17; 18 and 19 and 20 and 21 have phase plots of a circular form. This indicates that the two eigentriples represent a sine- and cosine pair with equal amplitude, frequency and thus represent a harmonic cycle. Naturally, these eigentriples will be included in the reconstruction in order to be able to represent the seasonal cycles.

Figure 4: Phase plot of the first 21 eigentriples after SSA decomposition steps.



Considering all the aforementioned the eigentriples $1, 2, \cdots, 21$ are chosen for the reconstruction and forecasting of the time series. The first eigentriple is slowly varying and does not form a sine-cosine pair with any of the other eigentriples, hence, this is the slowly varying trend component. Eigentriples $2, 3, \cdots, 21$ all form sine-cosine pairs, suggesting that these eigentriples make up the seasonal component of the time series. Separate reconstructions of the trend and seasonality components of the training set are shown below in Figures 5 and 6.

Figure 5: SSA reconstruction of the trend component of the training set.

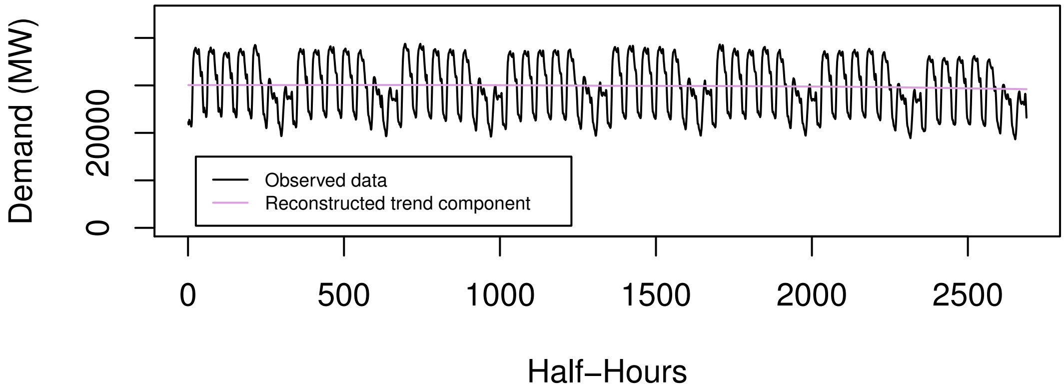
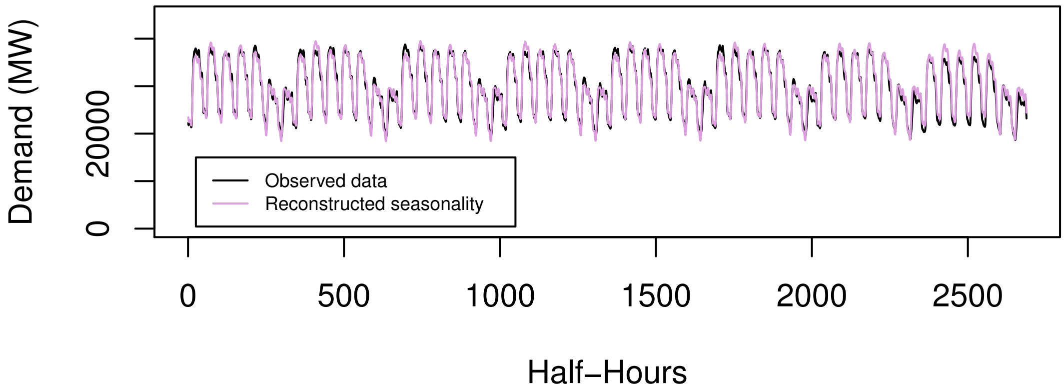
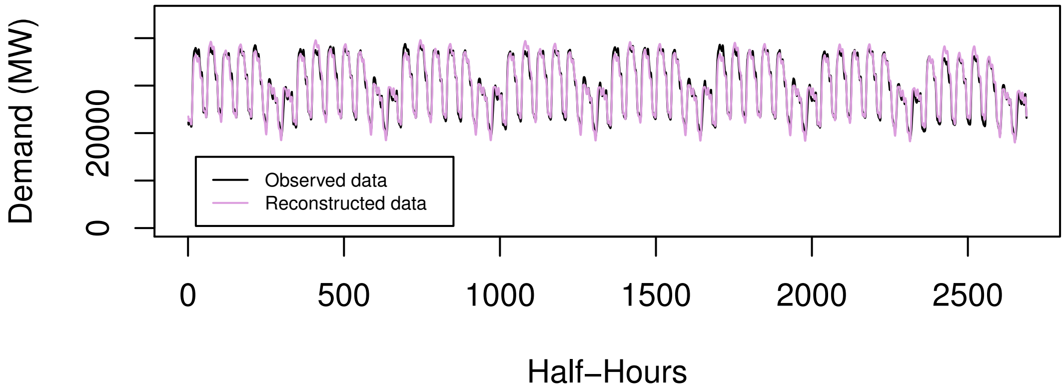


Figure 6: SSA reconstruction of the seasonal component of the training set.



As can be seen in the above figures, the trend component of the data is extremely slowly varying - almost too slow to distinguish with the bare eye - and the seasonality component makes up the largest part of the variation in data. This aligns well with the statements made in Section 3, where we stated that the smallest part of the time series variation can be attributed to the trend component, while the largest part of the variation was caused by the seasonal cycles. Lastly, the complete reconstruction of the training set is presented in Figure 7.

Figure 7: Plot of the observed data and the reconstructed data after SSA reconstruction steps.



The algorithm used to produce forecasts is described in Section 4.4. Besides, the accuracy of the forecasts produced using SSA is presented in Section 5.2.

4.3.4 Residual Autocorrelation Adjustment

To improve the accuracy of the forecasts made using SSA, an AR model is fitted to the residuals. After performing the decomposition and resconstruction steps training set, the errors of the reconstructed training set are calculated as follows for $t = 1, 2, \cdots, 2688$:

$$\hat{e}_t = y_t - \hat{y}_t. \tag{24}$$

To select the preferred order of the AR model we plot the sample autocorrelation function. After careful inspection of the number of autoregressive lags for which both the mean absolute percentage forecast errors and the SSE of a one- up to and including 48-step ahead forecast (starting from the observation at $t = 2688$) are minimised, an AR(4) model is chosen to model and predict the forecast errors. After a reconstruction of the training set an AR(4) model is fitted to the residuals of the reconstruction and used to predict the values of the residuals. These predicted values are added to the forecasts made using SSA in order to improve the forecasting accuracy. An ARMA model to model the residuals was considered as well; however, the addition of MA terms to the model did not improve the forecasting accuracy. The forecasting results for SSA recurrent forecasting with a fitted AR(4) model are presented in Figure 10 in Section 5.2.

4.4 Model Evaluation and Comparison

The models are evaluated in two different ways: firstly, the three different Holt-Winters formulations are compared to see if the forecasts made after the incorporation of an extra seasonal smoothing equation in the double seasonal Holt-Winters methods are more accurate than forecasts made merely using the standard Holt-Winters method. Secondly, the forecasting accuracy of the multiplicative double seasonal Holt-Winters method is compared to the well-specified benchmark model. The used measure of forecast accuracy is the mean absolute percentage error (MAPE), which is defined as:

$$MAPE = 100\frac{1}{N}\sum_{t=1}^{N}(\mid y_t - \hat{y}_t \mid / \mid y_t \mid), \tag{25}$$

where N denotes the number of observations for which forecasts are produced.

The procedure of determining the MAPEs for the h-step ahead ($h = 1, 2, \cdots, 48$) is as follows: First, parameters for the methods are estimated using the training set and afterwards, forecasts are made. Starting at observation Y_t ($t = 2688$), one- up to and including 48-step ahead forecasts are made for the subsequent 48 observations, $y_{t+1}, \cdots, y_{t+48}$. These forecasts, denoted by $\hat{y}_t(1)$, $\cdots$, $\hat{y}_t(48)$ are stored in the first column of a 48×1344 forecast matrix F. Afterwards, the observation Y_{t+1} is added to the set of known observations and forecasts are produced up to and including 48-steps ahead, starting from the observed value Y_{2689}. These forecasts are stored in the second column of F. This procedure is repeated until observation Y_{4031} is added to the set of known values and forecasts have been made. Each h-th row of F now contains the h-step ahead forecasts starting from observations Y_{2688} in column 1, and Y_{4031} in column 1344. To construct the MAPE for the h-step ahead forecast, the h-th row of F is selected. The forecasts for observations that are beyond the time span of our data (an example of this is the two-step ahead forecast $\hat{Y}_t(2)$, starting from $t = 4031$) are omitted from the set of forecasts. Lastly, the observations, y_t, corresponding to the forecasted values, $\hat{y}_t$, are selected from the data and the MAPE is calculated as described in (25).

The MAPE values of the different methods are compared to one another in Section 5.2 in order to determine which method performs best.

5 Results

5.1 Holt-Winters Exponential Smoothing: Parameter Estimation

The estimated values of the parameters for the three different Holt-Winters methods (Holt-Winters for within-day seasonality, Holt-Winters for within-week seasonality, double seasonal Holt-Winters) are presented in the Tables 1 and 2 below:

Table 1: Holt-Winters parameters, calculated from the training sample.

	Level α	Trend γ	Within-Day δ	Within-Week ω
Holt-Winters: within-day seasonality	0.99	0.86	1.00	–
Holt-Winters: within-week seasonality	0.83	0.00	–	1.00
Double seasonal Holt-Winters	0.88	0.00	0.83	1.00

Table 2: Holt-Winters parameters calculated from the training sample, where the method included an AR(1) model fitted to the residuals

	Level α	Trend γ	Within-Day δ	Within-Week ω	AR λ
Holt-Winters: within-day Seasonality	0.80	0.00	0.69	–	0.74
Holt-Winters: within-week Seasonality	0.01	0.00	–	0.42	0.92
Double seasonal Holt-Winters	0.01	0.00	0.18	0.33	0.93

Noteworthy about the results is that they imply that the fitting of an autocorrelation model to the residuals partly replaces the effect of the smoohting parameters. This effect is most clearly illustrated in the reduction of the level smoothing parameter α for all three Holt-Winters methods including an AR(1) model; however, the values of the estimates for α, γ, δ and ω reduce in all cases after the introduction of the AR(1) model, irrespective of the type of Holt-Winters exponential smoothing used.

5.2 Comparison of the Models

Below, the MAPE results for the three different Holt-Winters methods and the benchmark SARIMA model are shown in Figure 8, whereas the MAPE results for the three Holt-Winters methods including the autocorrelation adjustment can be found in Figure 9. The MAPEs for SSA with and without an AR model fitted to the residuals are presented in Figure 10 below.

Figure 8: MAPE results for the exponential smoothing methods with the results from the training period.

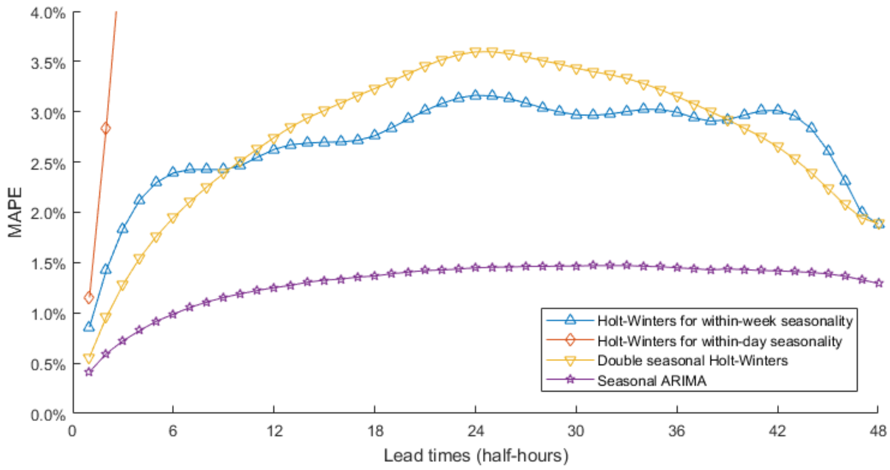


Figure 9: MAPE results for the exponential smoothing methods with the results from the training period and an included autocorrelation term.

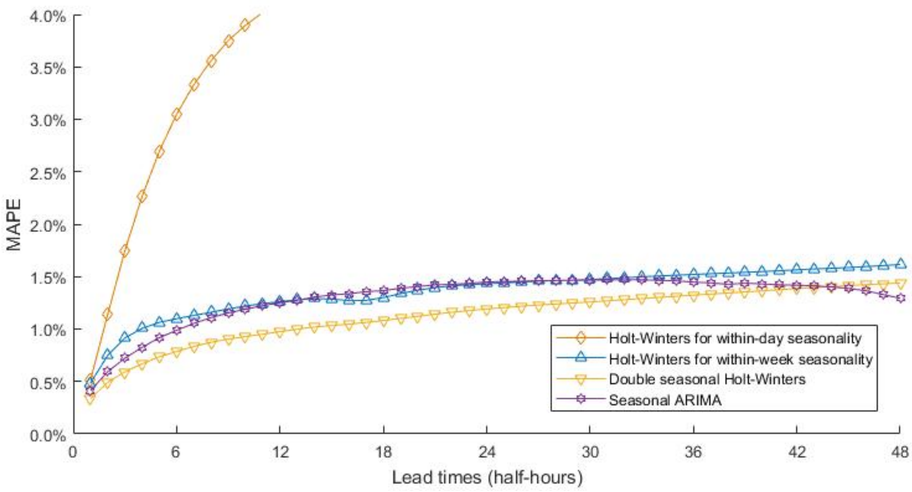
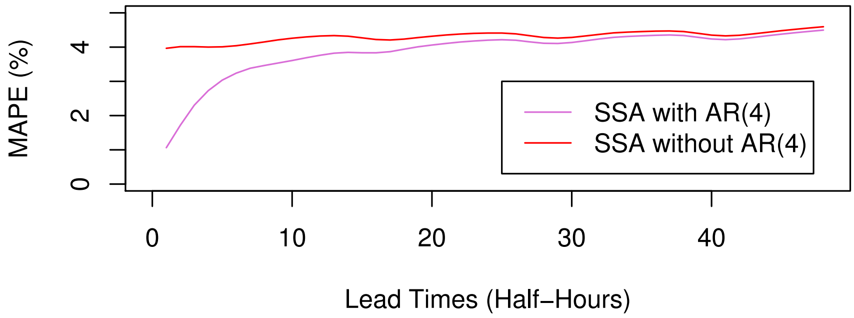


Figure 10: MAPE results for SSA with and without an AR(4) model fitted to the residuals.



Comparing the three different Holt-Winters methods, we note that the within-day Holt-Winters method is easily outperformed by both the benchmark SARIMA model and the other Holt-Winters methods with and without an autocorrelation adjustment. Possible reasons for that are discussed in Section 6.1. In contrast, the Holt-Winters methods accomodating either the within-week seasonal pattern or both seasonal cycles are more similar in their forecasting accuracy: for the models without autocorrelation adjustment, the models alternately outperform each other depending on the lead times of the forecasts, as can be seen in 8. However, as Figure 9 shows, the Holt-Winters model for within-week seasonality is outperformed by the double seasonal Holt-Winters method as well after the introduction of an AR(1) model for the residuals. The introduction of an autoregressive model to complement the Holt-Winters method is valuable for all three variations of the method, though: Figures 8 and 9 both show a decrease of the MAPEs after its introduction. On top of that, none of the Holt-Winters exponential smoothing methods outperforms the benchmark model before the inclusion of the AR(1) model; however after the inclusion of the AR(1) model, the double seasonal Holt-Winters method outperforms the benchmark model, while the the within-week Holt-Winters method is a good competitor of the benchmark SARIMA and achieves similar MAPEs.

Furthermore, Figure 10 shows that the SSA method without an autocorrelation method outperforms none of the other methods. The inclusion of an AR(4) model is a valuable addition to the model, especially regarding the forecast for short lead times. However, the SSA method with the inclusion of an AR(4) model for the residuals merely outperforms the Holt-Winters method for within-day seasonality, and does not succeed in outperforming any of the other proposed methods, including the benchmark model. Therefore, we conclude that of all methods evaluated, the forecasting performance of the double seasonal Holt-Winter method with autocorrelation adjustment is better than all other proposed methods.

6 Discussion

In Section 6.1 we will present a discussion of the methods applied during the research and try to give an explanation regarding some results that differ from what was expected beforehand. The section will be closed with suggestions for further research.

6.1 Discussion of the Applied Methods

As stated in Section 5.2, the Holt-Winters method to accommodate the within-day seasonal cycle performs very poorly in comparison with the other suggested methods. A possible explanation may be that the within-day seasonal cycle only constitutes a small part of the overall seasonality component, whereas the within-day seasonal cycle makes up the largest part of the seasonal component. This would explain why the MAPEs of the within-day Holt-Winters method are so large, and analogously, why the MAPEs of the within-week Holt-Winters method compete rather well with the double seasonal Holt-Winters method.

Secondly, we will touch upon the forecasting results of the SSA method. Considering the promising results achieved with SSA in the past, it was rather surprising to find that both the benchmark model, alongside with the within-week and double seasonal Holt-Winters methods, outperformed the SSA model. A possible reason for this could be that the selection of the eigentriples does not optimally separate the signal of the time series from its noise, and that thus; a relevant component of the time series is not taken into account when forecasting the time series, causing a reduced forecasting accuracy. A solution for this would be to consider more sophisticated grouping methods in order to be able to separate the independent components of the time series in a more satisfactory manner. Examples of such methods could be window length effects (Golyandina and Zhigljavsky, 2013).

Lastly, we will briefly discuss the fact that the replication of the MAPE results differ slightly from those presented in Taylor (2003). The most plausible explanation for this would be the sensitivity of the optimising method to the starting values. When implementing the Holt-Winters exponential smoothing method, we noted that a wide range of results could be achieved using different starting values, and although our reproduced results bear a fair amount of similarity those of Taylor, they are not exactly alike. Thus, a small difference in the starting values of the optimisation is likely to be the cause of this discrepancy.

6.2 Suggestions for Future Work

Although the double seasonal Holt-Winters method has certainly achieved promising results in our research, we suggest to investigate the external validity of the method by applying the method to different data containing a variety of different seasonal patterns. For example, in our data, the within-day seasonal cycle is nested in the within-week seasonal cycle. A topic for further research could be the forecasting performance of the double seasonal Holt-Winters method for data sets in which the seasonal cycles are not nested.

Although the intraday cycles are treated as being the same every day in our research this is not always the case. However, the double seasonal Holt-Winters cannot accommodate small variations in the seasonal cycles it captures. Hyndman et al. (2008) investigate the modelling of varying cyclical behaviour by means of an innovation state space approach with successful results. A possible approach to model varying cyclical behaviour within the double seasonal Holt-Winters method is the inclusion of more variables or dummy variables within the seasonal index equations of the Holt-Winters method.

As a last topic for further research we suggest the weighted combining of different methods that can accommodate multiple seasonalities. Caiado (2010) investigates this method by forecasting water demand in Spain and concludes that the combined forecasts are especially suitable for short-term forecasting.

7 Conclusion

Accurate and robust forecasting techniques are of vital importance to the National Grid; they are needed to be able to balance electricity demand and supply. The half-hourly data on the demand of electricity in England and Wales clearly shows two seasonal patterns: one within-day seasonal cycle of length 48 half-hours and one within-week seasonal cycle of length 336 half-hours. In this research the standard Holt-Winters exponential smoothing method was extended to accommodate two seasonal cycles, as opposed to one. An autoregressive model has been included in both the standard and the seasonal Holt-Winters methods to model the residuals and thus improve the forecast accuracy. Furthermore, the relatively new method of SSA was proposed to model and predict the time series. It was investigated whether (a) this extended double seasonal Holt-Winters method would provide more accurate forecasts than the standard Holt-Winters method, (b) the within-day, within-week and double seasonal Holt-Winters methods would outperform a more traditional SARIMA benchmark model and (c) whether forecasts made using SSA would outperform either one of the Holt-Winters methods or the SARIMA benchmark model. The within-day Holt-Winters method has a poor performance in comparison with other methods, whereas the within-week method is fairly competitive with the double seasonal Holt-Winters method and even outperforms the SSA method. However, we find that the double seasonal Holt-Winters method with an autoregressive model fitted to the residuals outperforms all other proposed methods, including SSA. This is in line with our first hypothesis regarding the forecasting performance of the double seasonal Holt-Winters method. The forecasts made using SSA, however, neither outperform the benchmark SARIMA nor the within-week and double seasonal Holt-Winters methods - although it provides fairly accurate forecasts for forecasts with very short lead times. Thus, our hypothesis regarding the singular spectrum analysis method is rejected. Finally, it can be concluded that the seasonal Holt-Winters is a very promising method of forecasting short-term electricity demand, either by itself or in combination with other forecasting techniques.

References

Abraham, A. and Baikunth, N. (2001). A neuro-fuzzy approach for modelling electricity demand in victoria. *Applied Soft Computing*, 1(2):127–138.

Abu-El-Magd, M. and Sinha, N. (1982). *Univariate and Multivariate Techniques For Short-Term Load Forecasting*. IFAC: New York City.

Alexandrov, T. (2009). A method of trend extraction using singular spectrum analysis. *REVSTAT - Statistical Journal*, 7(1):1–22.

Badran, I., El-Zayyat, and H., Halasa, G. (2008). Short-term and medium-term load forecasting for jordan's power system. *American Journal of Applied Sciences*, 5(7):763–768.

Box, G. and Jenkins, G. (1970). *Time Series Analysis: Forecasting and Control*. San Fransisco: Holden-Day.

Box, G., Jenkins, G., and Bacon, D. (1967). *Models for Forecasting Seasonal and Nonseasonal Time Series*. John Wiley and Sons: New York City.

Bunn, D. (1982). Short-term forecasting: A review of procedures in the electricity supply industry. *The Journal of the Operational Research Society*, 33(6):533–545.

Caiado, J. (2010). Performance of combined double seasonal univariate time series models for forecasting water demand. *Journal of Hydrologic Engineering*, 15(3):215–222.

Chatfield, C. (1978). The Holt-Winters forecasting procedure. *Journal of Applied Statistics*, 27(1):264–279.

Christiaanse, W. (1971). Short-term load forecasting using general exponential smoothing. *IEEE Trans Power Syst*, (2):900–902.

Cochrane, D. and Orcutt, G. (1949). Application of least square regression to relationships containing auto-correlated error terms. *Journal of the American Statistical Association*, 44(245):32–61.

Durbin, J. and Watson, G. (1950). Testing for serial correlation in least-square regression. *Biometrika*, 37(3-4):409–428.

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

Gilchrist, W. (1976). *Statistical Forecasting*. Wiley: Chichester.

Golyandina, N., Nekrutkin, V. V., and Zhigljavsky, A. (2001). *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman & Hall/CRC: Boca Raton, FL.

Golyandina, N. and Zhigljavsky, A. (2013). *Singular Spectrum Analysis for Time Series*. Springer: New York City.

Hassani, H. (2007). Singular spectrum analysis: Methodology and comparison. *Journal of Data Science*, 5(2):239–257.

Hassani, H. and Ghodsi, Z. (2015). A glance at the applications of singular spectrum analysis in gene expression data. *Biomolecular Detection and Quantification*, 4:17–21.

Holt, C. (2004). Forecasting seasonals and trends by exponentially weighted averages. *International Journal of Forecasting*, 20(1):5–10.

Hyde, O. and Hodnett, P. (1993). Rule-based procedures in short-term electricity load forecasting. *IMA Journal of Mathematics Applied in Business and Industry*, 5(1):131–141.

Hyndman, R. (2001). Box-Jenkins modelling. https://robjhyndman.com/papers/BoxJenkins.pdf.

Hyndman, R. and Athanasopoulos, A. (2001). *Forecasting: Principles and Practice*. https://www.otexts.org/fpp.

Hyndman, R., Gould, P., Koehler, A., Ord, J., Snyder, R., and Araghi, F. (2008). Forecasting time series with multiple seasonal patterns. *European Journal of Operational Research*, 191(1):207–220.

Hyndman, R., Makridakis, S., and Wheelwright, S. (1998). *Forecasting: Methods and Applications*. John Wiley and Sons.

Kalekar, P. (2004). Time series forecasting using Holt-Winters exponential smoothing. Unpublished. https://labs.omniti.com/people/jesus/papers/holtwinters.pdf.

Kirchgässner, G. and Wolters, J. (2007). *Introduction to Modern Time Series Analysis*. Springer: New York City.

Peng, C. and Ching-Wu, C. (2009). A comparison of univariate methods for forecasting container throughput volumes. *Mathematical and Computer Modelling*, 50(7-8):1045–1057.

Reid, D. (1975). A review of short-term projection techniques. In: Gordon ha (ed). *Practical Aspects of Forecasting*, pages 8–25.

Rodrguez-Aragn, L. and Zhigljavsky, A. (2010). Singular spectrum analysis for image processing. *Statistics and Its Interface*, 3:419–426.

Schoellhamer, D. (2001). Singular spectrum analysis for time series with missing data. *Geophysical Research Letters*, 28(16):3187–3190.

Shang, Q., Lin, C., Yang, Z., and Zhou, X. (2016). A hybrid short-term traffic flow prediction model based on singular spectrum analysis and kernel extreme learning machine. *PLOS ONE*, 11(8):Online publication.

Taylor, J. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing. *The Journal of the Operational Research Society*, 54(8):799–805.

Taylor, J. (2012). Short-term load forecasting with exponentially weighted methods. *IEEE Transactions On Power Systems*, 27(1):458–464.

Winters, P. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6(3):324–342.

A Values of the BIC Criterion for the SARIMA Model

Table 3: Values of the BIC criterion for various SARIMA models, with the order of differencing set to 0.

Number of MA lags/ Number of AR lags	1	2	3	4
1	3.5460	3.5515	3.5522	3.5516
2	3.5526	3.7613	3.7649	3.7624
3	3.7378	3.8210	3.8816	3.8777
4	3.7262	3.7673	3.9300	3.9501

Table 4: Values of the BIC criterion for various SARIMA models, with the order of differencing set to 1.

Number of MA lags/ Number of AR lags	1	2	3	4
1	3.5524	3.5527	3.5526	3.5522
2	3.5528	3.5523	3.5572	3.5564
3	3.5529	3.5573	3.5562	3.5563
4	3.5525	3.5566	3.5571	3.5571

B Holt-Winters Exponential Smoothing Autocorrelation Plots

Figure 11: Sample autocorrelation plot of the standard Holt-Winters method capturing the within-day seasonal cycle performed on the training sample.

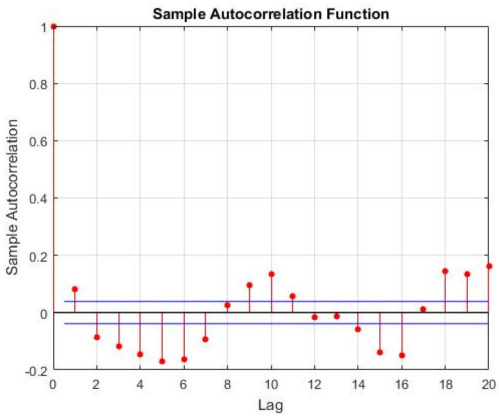


Figure 12: Sample autocorrelation plot of the standard Holt-Winters method capturing the within-week seasonal cycle performed on the training sample.

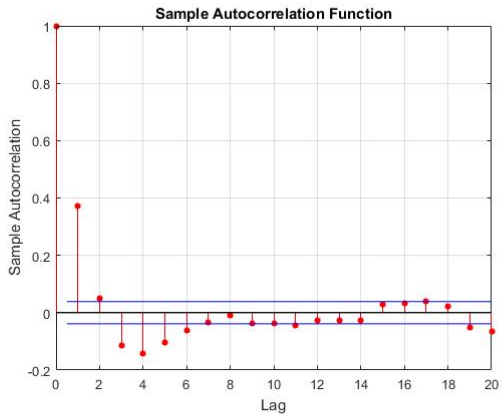


Figure 13: Sample autocorrelation plot of the double seasonal Holt-Winters method performed on the training sample.