Short-term electricity demand forecasting in the time domain and in the frequency domain

Abstract

This paper compares the forecast accuracy of different models that explicitely accomodate seasonalities in the time domain with methods in the frequency domain. The forecast accuarcy of the methods are evaluated by their MAPE's using a time series of half-hourly electricity demand for England and Wales. The methods in the time domain include a well-specified multiplicative seasonal ARIMA model, standard Holt-Winters method, double seasonal Holt-Winters method both with and without residual auto-correlation correction. The methods in the frequency domain introduced in this paper are Fourier series that are modified to accomodate stochastic dependence. The last class of models in the frequency domain is shown to perform almost as good as the methods in the time domain and is promising in terms of accuracy and adaptability.

Jonathan van de Ruitenbeek	Student (id 412442)
Mikhail Zhelonkin	Supervisor
Michel van de Velden	Co-Reader

Erasmus University Rotterdam, Erasmus School of Economics

Econometrics and Operations Research

Bachelor Thesis, July 2, 2017

1 Introduction

Demand forecasting is thoroughly discussed in the Econometric literature and at the same time of great importance for business and has many practical applications. Over the past century, scholars have developed a variety of sophisticated models that can be used for modelling time series with many different features. The interest of this paper lies in predicting half-hour ahead electricity demand for the National Grid. The National Grid is an international electricity and gas company that is responsible for electricity transmission in England and Wales. They play a crucial role in connecting millions of citizens efficiently to the energy they consume. Electricity demand forecasts are of crucial importance to balance the supply and demand between generators and suppliers, and accurate forecasts affect multiple levels of production systems. Accurate forecasts can lead to lower costs, increasing competitiveness and efficient management. While there is a wide range of forecasting methods available for use, each method has it's own advantages and disadvantages.

In this paper I aim to forecast a time series of half-hourly electricity demand in England and Wales for a fortnight in June 2000. Given the strong seasonal patterns in the time series I focus on univariate methods from the time domain and the frequency domain that incorporate seasonalities. Only univariate methods are considered because of the short lead times. For longer lead times multivariate models might be more suitable. An example of multivariate methods applied to forecasting electricity demand is given by Taylor & Buizza (2003) which discusses electricity demand forecasts based on a model that uses weather variables as input. but For this application however, multivariate models would be too demanding in terms of half-hourly input.

Of particular interest is the paper of Taylor (2003), where he compares the multiplicative double seasonal ARIMA model with several variants of the Holt-Winters model, by forecasting short-term electricity demand. The most important contribution of his paper lies in developing a double seasonal Holt-Winters model which is able to account for more than one seasonal pattern. The methods discussed in Taylor (2003) from the time domain are compared with three other forecasting methods from the frequency domain.

Modelling time series with trigonometric functions is less widely done in practise than other methods, but is very appealing given the features of the time series of electricity demand. The time series contains two strong seasonal patterns, which can be modelled very well with trigonometric functions. The models discussed in this paper share some convenient features that I will discuss in section 3.3. The methods in this paper are based on the Fourier series and as far as I know, Fourier series have not yet been use for forecasting short-term electricity demand. The first model fit to the data is a full Fourier series model, the second method considers a adaption of the first method by correcting for residual auto-correlation and the third model uses a combination of an ARIMA model for modelling short-term variation and Fourier terms for modelling multiple seasonal patterns. Cyclical patterns are analyzed by means of spectral analysis. The outline of the paper is as follows. In the next section I highlight some of the univariate methods discussed in the literature for modelling electricity demand time series and literature on Fourier analysis. In the section that follows I discuss the models used in this paper in more detail. I then present and discuss the results and end with the final section that contains a summary and a conclusion.

2 Literature Review

A wide range of univariate methods for modelling time series is discussed in literature. Many of these methods have been applied to forecasting short-term electricity demand. Taylor et al. (2006) for example, discusses the accuracy of six univariate methods for forecasting electricity for short lead times. They conclude that simple exponential smoothing methods which are easy and quick to implement outperform more complex alternatives such as the Principal Components Analysis and neural networks. Exponential smoothing methods are based on a pragmatic approach to forecasting and are very popular and widely used for short-term forecasting. The different components of the model are are intuitive to use, computational effort is kept low and a suprising accuracy can be achieved with minimal model identification. Makridakis et al. (1979) and Makridakis et al. (1982) found little difference in large scale studies between the accuracy of exponential smoohting methods and well-specified ARIMA models. Many different variants of exponential smoothing methods have been developed and a detailed overview of the different formulations are given by Gardner (1985) and Gardner (2006). A noticeable development was made by Taylor (2003), who discusses an extension to the standard Holt-Winters method. He considers an exponential smoothing formulation that accounts for double seasonality, by introducing an extra seasonal index compared to the standard Holt-Winters method and adjusting the forecasts and updating equations accordingly. The double seasonal Holt-Winters method can be extended for any number of seasonalities, though in practise one will often not find more than two or three seasonal patterns. Another very popular approach for modelling time series is the ARIMA model with all of it's variants. The method is extremely popular and since Box and Jenkins (1976) discuss a systematic method for fitting an ARIMA model it has been very widely used. The ARIMA model can incorporate different features of time series such as a trend and seasonality and it often serves as a benchmark for comparing other models. An interesting extension of the seasonal ARIMA model is the double seasonal ARIMA model which is discussed by Box et al. (2015). They argue that the single seasonal ARIMA model can be extended in such a way that it can account for multiple seasonal patterns in the data.
A less widely used approach for modelling time series is given by the use of trigonometric functions for modelling seasonal patterns. This area of research was initiated by the research of Joseph Fourier, who was a French mathematician and physicist who is very well known by initiating research on the Fourier series. Through Fourier's research the fact was established that an arbitrary function can be fully represented by a trigonometric series. The original motivation of Fourier series was to model heatwaves, but later scholars found that Fourier series analysis can be applied to

a wide range of topics and now application of Fourier series analysis has found it's way in time series analysis as well. There are a variety of textbooks such as the book of Bloomfield (2004) which discuss the application of Fourier series to modeling time series. De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011), suggest the Fourier series approach for modelling time series that contain multiple seasonal patterns. They propose an innovations state space modeling framework for modelling time series with multiple and complex seasonal patterns. The proposed model is based on Fourier series and incorporates Box-Cox transformations and ARMA error-correction. McLoughlin et al. (2013) discuss the application of Fourier transform in forecasting individual demand and conclude that the method performs well. González-Romera et al. (2008) discuss the application of forecasting monthly electricity demand by modeling the seasonal patterns by Fourier series and modeling the trend by a neural network. In this paper different trigonometric models are developed and fit to the data and shown to be almost as good as very popular methods, along with the fact that the models have some very convenient features.

3 Methodology

3.1 Data

The data used in this paper is a time series of half-hourly electricity demand in England and Wales over a period of 12 weeks, starting from Monday 5 June 2000 to Sunday 27 August 2000[1]. The complete 12-week sample consists of 4032 observations, where the sample is split in an estimation sample of 2688 observations and a validation sample of 1344 observations. All models are fit on the estimation sample and evaluated by forecasting with a forecast horizon of 48 half-hour periods. The next figure displays the Seasonal-Trend Decomposition Based on Loess, designed by Cleveland et al. (1990) for the estimation sample. The STL-decomposition is a computationally efficient filtering procedure that decomposes a time series into a seasonal, trend and remainder components. The method is developed by Cleveland et al. (1990) using the Loess estimation method, which is a method for estimating non-linear relationships. The decomposition procedure facilitates the analysis of the time series by disaggregating the series into feature-based sub-series, which provides us a graphical overview of the different features of the time series. The next figure shows the STL-decomposition of electricity demand.

[1]The data can be found in the forecast package in R, which is developed by Hyndman (2007)

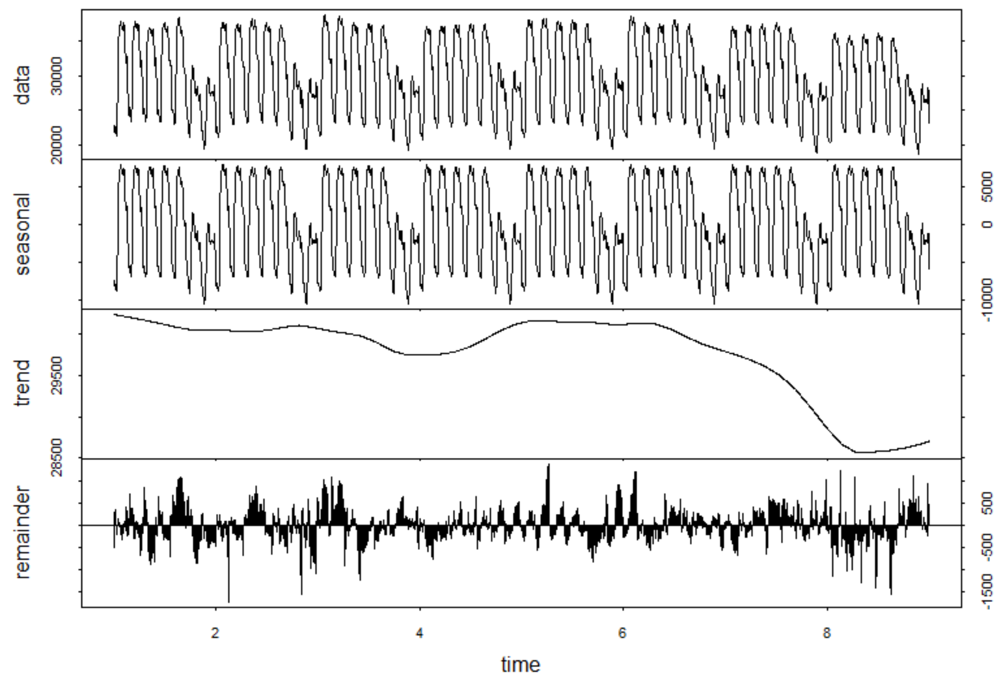

Figure 1: STL-decomposition based on Loess

The numbers on the x-axis denote the weeks of the estimation sample. The upper part of the graph is a plot of the raw data set. The next part of the graph contains a plot of the seasonal pattern that is extracted from the data. The graph that follows contains a plot of the estimated trend from the data and in the last plot the remainder of the data which could not be captured by trend seasonality is displayed. It can be observed that the time series contains a very strong weekly 336-period seasonal pattern and a strong daily 48-period seasonal pattern. Due to the strong seasonality, the mean of the observations is not constant anymore, but instead moves according to a cyclical pattern. The seasonal patterns seems to evolve stable and constant over time. Because the time series exhibits two strong seasonal patterns, there is strong appeal to models that can capture this features and therefore all of the models discussed in this paper can accommodate one or multiple seasonal patterns. The trend is slightly downwards-sloping for the largest part of the data, while in the end the trend is estimated upward-sloping. The last part of the graph contains the remainder of the series, which is the part of the time series that can not be captured by trend or seasonality. Performing an Augmented Dickey-Fuller test on the estimation sample rejects the null hypothesis of a unit root against the trend-stationary alternative, indicating that the data is trend-stationary.

In the next section I give an overview of the methods used by Taylor (2003) which deal with multiple seasonality and afterwards I discuss the models based on the Fourier series.

3.2 Methods by Taylor (2003)

In this paper I implement the methods discussed by Taylor (2003) to forecast the electricity demand, which are discussed in this section. The methods are given by the standard Holt-Winters method accounting for multiplicative seasonality at 48 half-hourly periods, the standard Holt-Winters method accounting for multiplicative seasonality at 336 half-hourly periods, the double seasonal Holt-Winters method accounting for both seasonalities and the multiplicative seasonal ARIMA model. k-step ahead forecasts are computed for k=1,...,48 and are evaluated by their MAPE's.

3.2.1 Exponential smoothing methods

Exponential smoothing methods are very popular for various reasons, which are discussed in the literature review. It is a popular and widely used scheme to produce a smoothed time series that is related to the idea of (weighted) moving averages. Instead of weighting past and recent observations equally, exponential smoothing assigns more weight to recent observations compared to past observations in forecasting future observations. As such, observations are assigned exponentially decreasing weights as the observations get older. Hence, the name 'Exponential smoothing', which is given to the method because when time passes by, the smoothed level becomes a weighted average of more and more past observations. Exponential smoothing methods come in different forms, and in this paper I implement the methods discussed by Taylor (2003). As noted earlier, the electricity demand time series contains strong seasonal patterns, which indicate that the Holt-Winters model with trend and seasonality are most suitable for modeling the time series. The models are implemented under the assumption of an additive trend and multiplicative seasonality. The trend is assumed additive in the sense that the trend is added to the underlying level of the series. The seasonality is multiplicative in the sense that the underlying level and trend are multiplied by the seasonal index. Other formulations are given by different combinations of additive or multiplicative seasonality and additive or multiplicative trend and damped trend or not. The multiplicative formulation for seasonality is more widely used than the additive model formulation and simplicity I stick to the methods discussed by Taylor (2003).
The standard Holt-Winters method is suitable for modelling time series with a trend and one seasonal pattern. The method has an additive version and a multiplicative version, where in this paper the version with additive trend and multiplicative seasonality is implemented for daily (seasonality = 48) and weekly (seasonality = 336) seasonality. The multiplicative formulation of the model is given by the following set of equations:

$$\text{Level}: \quad S_t = \alpha\Big(\frac{y_t}{I_{t-s}}\Big) + (1-\alpha)(S_{t-1} + T_{t-1}) \tag{1}$$

$$\text{Trend}: \quad T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1} \tag{2}$$

$$\text{Seasonality}: \quad I_t = \delta\Big(\frac{y_t}{S_t}\Big) + (1-\delta)I_{t-s} \tag{3}$$

$$\text{Forecast}: \quad y_{t+k} = (S_t + kT_t)I_{t-s+k} \tag{4}$$

where the level S_t, trend T_t and seasonality I_t are recursively updated every iteration and k-step ahead forecasts can be made by computing y_{t+k}. The smoothing parameters are given by α, γ and δ. The length of the season is given by s, t denotes the index of the current iteration and k denotes the forecast horizon. Initialization of the standard Holt-Winters models is done by a regression-based technique recommended by Hyndman et al. (2008). The initial level, trend and seasonality are initialized by classical decomposition of the time series into seasonal, trend and irregular components, based on a multiplicative model. The model is multiplicative in the sense that the time series is assumed to consist of the multiplication of the trend component, the seasonal component and the random component. The trend values for the first two seasons are determined by using a centered moving average, i.e. symmetric window with equal weights with the length of the season. The trend is regressed on the variable t which consists of the values 1,2,3... and the coefficient corresponding to the variable t is set as the initialization of the trend, while the constant in the regression is set as the initialization of the level. The seasonal figure is computed by taking the ratio of the actual observation to the average of the first period.

The double seasonal Holt-Winters method, first introduced by Taylor (2003), is able to account for two seasonal patterns in a time series. The double-seasonal formulation is similar to the standard Holt-Winters formulation, but additionally a new seasonal index is introduced. The formulation for additive trend and multiplicative seasonality is given by the following set of equations, formally denoted as follows:

$$\text{Level}: \quad S_t = \alpha\Big(\frac{y_t}{D_{t-s_1}W_{t-S_2}}\Big) + (1-\alpha)(S_{t-1}+T_{t-1}) \tag{5}$$

$$\text{Trend}: \quad T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1} \tag{6}$$

$$\text{Seasonality 1}: \quad D_t = \delta\Big(\frac{y_t}{S_t W_{t-s_2}}\Big) + (1-\delta)D_{t-s_1} \tag{7}$$

$$\text{Seasonality 2}: \quad W_t = \omega\Big(\frac{y_t}{S_t D_{t-s_1}}\Big) + (1-\omega)W_{t-s_2} \tag{8}$$

$$\text{Forecast}: \quad y_{t+k} = (S_t + kT_t)D_{t-s_1+k}W_{t-s_2+k} \tag{9}$$

where compared to the standard Holt-Winters formulation an extra equation for seasonality is introduced, the equations for seasonality are adjusted to for the other seasonality and the forecasts are adjusted for both seasonalities. The new seasonal index is given by W_t and an extra smoothing parameter is given by ω.
Initialization of the double seasonal Holt-Winters method is done by the method described in Taylor (2003) and Williams and Miller (1999). The initial trend is initialized as the average of (1) and (2) where (1) is calculated as 1/336 of the difference between the mean of the first 336 observations and the mean of the second 336 observations. (2) is calculated as the average of the first differences for the first 336 observations. The level component is initialized as the mean of the 672 observations minus 336.5 times the initial trend. Initial seasonal values are computed for every period in the season. The s_1 initial seasonal values for within-day seasonality D_t are calculated as the average of the ratios of the actual observations to the 48-point

centred moving average from that particular period in the season. The observations in the first week are used for initialization for the within-day seasonal indices. The s_2 initial seasonal values for within-week seasonality W_t are calculated in the same way as the daily-seasonal values, and then divided by the corresponding within-day seasonal index D_t. Note that the values of the level and trend are considered as observations at $t = 0$ and initial values of the seasonal indices are considered as observations corresponding to time $t - m$ where m corresponds to the seasonal period and $t = 0$.
The models are estimated in Matlab by minimizing sum of the squared 1-step ahead forecast errors for the estimation sample, for which I used the first 8 weeks of the data set in a similar manner as Taylor. The last 4 weeks of the time series is used for post-sample forecasting evaluation. This implies that the estimation sample is given by the first 2688 half-hourly observations and the validation sample is given by the remaining 1344 half-hourly observations.
After applying the exponential smoothed methods discussed above, inspection of the forecasts errors displays significant first-order autocorrelation. The standard and double seasonal Holt-Winters forecasts are adjusted by fitting an AR(1) term to the error terms, which was initially proposed by Reid (1975) and Gilchrist (1976), who found the same phenomenon. The adjustment is done by fitting an AR(1) model to the 1-step forecast errors, formally denoted as follows:

$$e_t = \lambda e_{t-1} + \xi_t \tag{10}$$

such that the forecasts are adjusted as follows:

$$\hat{y}_{t+k} = \hat{y}_{t+k} + \phi^k (y_t - \hat{y}_t) \tag{11}$$

where $\hat{y}_{t+k}$ denotes the k-step ahead forecast obtained by the Holt-Winters procedure. Reid (1975) and Gilchrist (1976) adjust the forecasts in a two-stage estimation, but Chatfield (1978) suggests a single-stage estimation approach for incorporating the error term correction, by simultaneously estimating the model parameters and ϕ by minimizing the sum of the squared 1-step-ahead errors. The latter procedure is applied in this paper.

3.2.2 Multiplicative seasonal ARIMA

The autoregressive integrated moving average (ARIMA) is a very popular and well-known method to model time series that involve autoregressive dependence, moving average dependence, and differencing operations. The seasonal ARIMA model involves backshifts of the seasonal periods, which allows for modelling seasonality. The multiplicative seasonal ARIMA model can be written as follows:

$$\phi_p(L)\Phi_P(L^s)\nabla^d\nabla_s^D y_t = c + \theta_q(L)\Theta_Q(L^s)z_t, \tag{12}$$

where y_t denotes demand in period t, z_t denotes the error term, c is a constant term, s is the number of periods in the seasonal cycle, ∇ denote the difference operator, d

denotes the order of difference, D denotes the order of seasonal difference, s denotes the seasonal period, ϕ_p,Φ_P, θ_q and Θ_Q, are polynomial functions of orders p, P, q and Q respectively. The model is multiplicative in the sense that the non-seasonal factors are multiplied by the seasonal differenced factor. The model is often expressed in the form ARIMA$(p, d, q) \times (P, D, Q)_s$. Application of the model to the time series of electricity demand yields the value $s = 336$ for within-week seasonality, $D = 1$ to account for non-stationarity for observations made in the same period in successive weeks. After carefully checking the auto-correlation plot and the partial-autocorrelation plot, a model was chosen with an intercept, an AR(2), a seasonal AR(2) and a seasonal MA(2) term. Other models were fit as well, but this model with hourly-dependence performs best.

3.3 Spectral Analysis

3.3.1 Spectral analysis

The analysis of time series can basically be carried out from two approaches. The first approach being from the time domain perspective and the other approach being from the frequency domain perspective. All models discussed above are based on the time domain approach, while spectral analysis is approached from the frequency domain. Spectral analysis applied to economic time series is a relatively not very widely used technique that attempts to explain the underlying pattern of a given time series by sine and cosine waves. In spectral analysis, the time series is broken down into it's cyclical components which are linked to the amount of variation they explain. As our data set contains very strong cyclical patters, there is a strong appeal to analyze the time series through spectral analysis. Before going into the application of spectral analysis to forecasting our time series, it is useful to review spectral analysis.
In their paper 'Spectral analysis in business forecasting', Chan and Hayya (1976) explain the concept and application of spectral analysis for non-mathematicians. 'Spectral analysis' comes from physics and is about analysis of light, sound, or other system of vibrations into a spectrum. Given a data set that contains fluctuations, one can define the fluctuations as noise. When this noise is put through a filter or window, in a similar manner as a physicist puts white light through prism, the noise can be broken down into components, which is called a 'spectrum'. For a given time series, the cyclical patterns can be broken down into a weighted combination of sine and cosine functions at different frequencies. The variance is broken down in such a way that the proportion of explained variance is associated with the harmonics at different frequencies. Spectral analysis is similar to analysis of variance, i.e. the larger the spectrum at a particular frequency, the larger is its contribution to the variance. The spectrum breaks down the sample variance of a time series into different components by associating these components with different frequencies. It can be shown that the spectra are a decomposition of the variance and as such the relative importance of frequency k/p_i can be identified, see for example Rayner (1971). It can be shown that there is a one-to-one relation between the spectrum of a time series and the autocovariance function, therefore the same information is depicted by both, but

with a slightly different interpretation. Spectrum analysis has an interpretation of analysis of variance. The field of this research is also closely related to harmonic analysis or Fourier analysis.

In this paper I discuss three models based on the Fourier series, which are analyzed and identified by means of spectral analysis. The first model fitted is a full Fourier model. As we will see, this model is a deterministic function of time and does not include any stochastic dependence. The next models consider an adaptive version of the first model, such that the models include stochastic dependence on previous observations. The second model is an adaptive version of this model which considers residual auto-correlation correction and the third model consists of a combination of trigonometric functions and an ARMA model.

Full Fourier model

The first model based on trigonometric series is the full Fourier model. Fourier (1808) proved that an arbitrary set of data points can be completely represented by a trigonometric series, which implies that a given time series can be represented by a weighted combination of sine and cosine waves. All methods of fitting sinusoids to time series fall under the wider term of Fourier analysis. One of the basic results of Fourier analysis is that an arbitrary analytic function defined over a finite interval, can be represented by a weighted sum of trigonometric functions of harmonically increasing frequencies to any desired degree of accuracy. If K is chosen as N/2, where N denotes the sample size, the model perfectly fits the data, see e.g. Chan and Hayya (1976). The full Fourier series model is given by the following equation:

$$y_t = c + \sum_{k=1}^{K} \alpha_k \sin\Big(\frac{2\pi kt}{N}\Big) + \beta_k \cos\Big(\frac{2\pi kt}{N}\Big) \tag{13}$$

where y_t denotes observation of the time series at time t, c denotes a constant, α_k and β_k denote coefficients that weight the different harmonics, K is given by $N/2$ where N denotes the estimation sample size. The fact that K is expressed by the number N/2 contributes to the aliasing problem and the number N/2 is also called Nyquist folding frequency. Nyquist proved that no power from frequencies greater than $N/2$ contribute to the spectral density, which is among others explained by Koopmans (1995). The parameters are estimated by least squares. The residuals of the estimation are zero, because by choosing K=N/2 the Fourier series will perfectly fit the data. Forecasts are based on the estimation sample. Forecasts obtained by this model assume that the historical pattern exactly repeats itself, without any stochastic dependence on previous observations. Note that the forecast of observation i is the same for all k-step ahead forecasts, i.e. the 1-step ahead forecast of observation i is the same as the 48-step ahead forecast of observation i, therefore the MAPE's will the same for $k = 1,...,h$, where h denotes the forecast horizon.

Partial Fourier model with residual auto-correlation correction

The next model applied to the data set considers an adaptation of the full Fourier model. The full Fourier model is adapted such a way that it includes a smaller number of harmonics so that there are residuals left which are then modelled by a simple ARMA model. The model is adapted such that it can accommodate multiple seasonalities, by writing the current observation as a function of sine and cosine waves at the most important frequencies, i.e. including cycles of period kp_i. The model is formally denoted as follows:

$$y_t = c + \sum_{i=1}^{M} \sum_{k=1}^{K_i} \alpha_{i,k} \sin\Big(\frac{2\pi kt}{p_i}\Big) + \beta_{i,k} \cos\Big(\frac{2\pi kt}{p_i}\Big) + e_t \tag{14}$$

where electricity demand y_t is modelled by a constant and some Fourier terms. In this model c denotes a constant, M denotes the number of periods which are explicitly modelled, K_i denotes the degree of complexity for period p_i, $\alpha_{i,k}$ and $\beta i, k$ denote the weight coefficients of the different harmonics. Including an extra seasonality is easily done by adding an extra period p_i and choosing the degree of complexity K_i. The Fourier series is a weighted combination of multiple periodic functions, which repeats itself every p_i periods. Please note that the trigonometric part of the model is a simplification of the model described by equation (13), because only the most dominant periodic cycles are taken into account.

To find the most important harmonic terms at the frequencies that contribute most to the variance of the series, a periodogram is constructed. A periodogram is used to identify the most important oscillation patterns and is given by the spectrum of the series plotted against it's corresponding frequencies. Time series data are measured at discrete, fixed intervals t in time at a frequency of $1/t$. The transformation of the signal from the time domain to the frequency domain can be done by the Fourier Transform, which transform a signal from the time domain to a signal in the frequency domain by summing up different sinusoids. The signal in the frequency domain is described by it's amplitude, frequency and it's phase t. At this point, the amplitude (maximum range of variation) and phase of the signal can be plotted at every frequency (or across the entire spectrum) which gives the frequency domain representation of the signal (spectral density). This information can be very useful in identifying periodicity in the series and is represented in a periodogram.

The periodogram is constructed in R where it is calculated using a Fast Fourier transform, and smoothed with a series of modified Daniell smoothers (moving averages giving half weight to the end values) to get a consistent estimator of the spectrum. The Fast Fourier transform is an algorithm for computing the discrete Fourier transform of a data series at all of the Fourier frequencies, in a computational efficient way. It is worth mentioning that the Fourier model and the Fast Fourier transform are two different things which are not directly related to each other. Since a very strong 48-period and 336-period seasonal pattern is observed, two periodograms are constructed for the ease of interpretation. In the first periodogram a frequency of 1 corresponds to 48 half-hourly periods and in the second periodogram a frequency of 1 corresponds to 336 half-hourly periods.

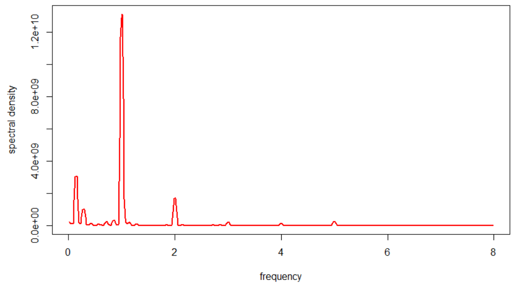

Figure 2: Periodogram at daily periods

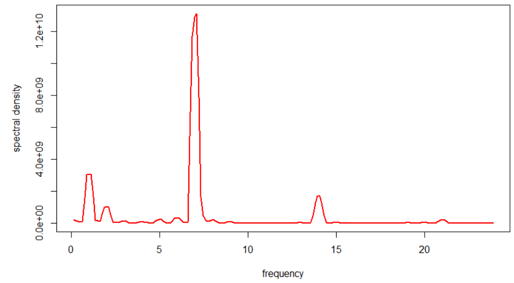

Figure 3: Periodogram at weekly periods

From the first periodogram it can be observed that a dominant peak is located at the frequency of 1/48, and other minor frequencies are located at multiples 2...5 of this frequency. From the second periodogram it can be observed that a dominant peak is located at the frequency of 1/336 and zoomed in it can be observed that minor peaks are located at multiples 2...5 of this frequency.
The model is now constructed as follows. I include two periods, being the daily period $p_1 = 48$ and the weekly seasonal cycle p_2 =336. Based on the information from the periodogram, I chose $K_1 = 5$ and $K_2 = 10$, such that the dominant cyclical patterns are modelled by harmonic terms at frequencies 1/48 and the 5 multiples thereof, and at frequencies 1/336 and 10 multiples thereof. Please note that the harmonic term corresponding to p_2 and $K = 7$ is not included in the model, because that harmonic term is essentially the same as the harmonic term corresponding to p_1 and $K = 1$. After fitting the Fourier model to the data, a model is chosen for the residuals based on the (partial) auto-correlation plot of the residuals. After careful inspections of the residuals, it is observed that the residuals display both a strong short-term half-hourly dependence and a strong longer-term dependence on the day before. The observed relationship is modelled by fitting an AR(1) term and an AR(48) to the residuals, formally denoted as follows:

$$e_t = \phi_1 e_{t-1} + \phi_2 e_{t-48} + \eta_t \tag{15}$$

where e_t denotes the residuals of the partial Fourier model for the time series of electricity demand, e_{t-1} and e_{t-48} denote the lags of residuals, and η_t the error term of the regression. The forecasts from the partial Fourier model are corrected by adding the expected forecast-error to the forecasts.

Partial Fourier representation combined with an ARMA model

The last model discussed in this paper also considers an adaptive version of the full Fourier model. The model is fitted with a number harmonic terms at the frequencies that explain most of the variance together with ARMA terms to explain the remaining variation of the series. The idea of a model that is estimated based where the data is modelled by a combination of trigonometric functions and an ARMA model is discussed by De Livera, Hyndman, & Snyder (2011) in the state-space modelling framework, but in this paper the model is analyzed and estimated within the conventional framework. The model is formally denoted as follows:

$$y_t = c + \sum_{i=1}^{M} \sum_{k=1}^{K_i} \alpha_{i,k} \sin\Big(\frac{2\pi kt}{p_i}\Big) + \beta_{i,k} \cos\Big(\frac{2\pi kt}{p_i}\Big) + N_t \tag{16}$$

where the seasonal cycles of the electricity demand y_t are modelled by a constant, the Fourier terms and where the remainder which is modelled as an ARMA process. Interpretation of parameters and coefficients is the same as the previous model and periodicity is identified in the same way as the previous model. Based on the periodogram, the number of periods is given by $M = 2$, where period $p_1 = 48$ and period $p_2 = 336$. The number of Fourier terms K_1 is chosen equal to 5 and K_2 equal to 10. The ARMA model is chosen based on carefully analyzing the (partial) auto-correlation plot and comparing the in-sample forecasting performance of different models. Short-term lags are added at significant low-order lags to account for short-term variation and long-term lags are added to account for long-term variation. The final model chosen for N_t is denoted as follows:

$$N_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \phi_4 y_{t-48} + \theta_1 z_{t-1} + \theta_2 z_{t-2} + \theta_3 z_{t-3} + \eta_t \tag{17}$$

where N_t denotes the ARMA process, y_{t-i} denotes the i-period lag of y_t with ϕ_i the corresponding coefficient, z_{t-i} denotes the i-period lag of the residual at time t with θ_i the corresponding coefficient and η_t denotes the error term of the regression. The full model is estimated in one stage estimation process, more specifically as an ARIMAX where the harmonics are added as external regressors.

With the models described above one can easily model multiple seasonalities of any length by adding multiple Fourier terms of different frequencies. Smooth seasonality is accounted for by small K, but more complex seasonal patterns can be modelled by a larger number for K. The model allows for non-integer seasonal periods, such as modeling the gap year, where the seasonal period is often set to 365.25 in case of daily seasonality. The data used in this paper does not contain any days with unusual demand, such as holidays. Taylor mentions that demands on these days is usually hard to forecast with univariate methods and that therefore in practise interactive facilities that override the system are used. However, the model decribed above can deal with unusual demand by including dummies for 'special' days such as Easter and Chistmas. Given the attractive features of the model, there is a strong appeal for applying the model to forecast electricity demand. The model accounts for daily seasonal cycle and the within-week seasonal cycle and, where the trigonometric part of the model assumes fixed seasonal patterns that do not change over time, however, by adding sesaonal lags of y_t the model accounts for seasonal variation over time. All of the above models are evaluated by their MAPE's of k-step ahead forecasts where k=1,..48, based on a rolling forecast origin. The models are updated by adding new observations, but parameters of the models described in this section are assumed to stay constant over time and are not re-estimated, since the process of model identification is impractical in the application of online demand forecasting.

4 Results

This section discusses the results for all of the models discussed in this paper. First, an overview is given of the results of the methods by Taylor (2003), then the results for the other models are presented and discussed and compared. Section 4.1 discusses the parameter estimates of the exponential smoothing methods, the parameter estimates of the multiplicative seasonal ARIMA and the estimation results of the Fourier models. In section 4.2 I compare and discuss the MAPE's of the different models and finally an evaluation of the models is presented.

4.1 Estimation results

The estimated smoothing parameters of the Holt-Winters models are presented in table (1) and the estimated smoothing parameters for the Holt-Winters models with residual auto-correlation correction are displayed in table (2).

Table 1: Results Holt-Winters parameters without auto-correlation adjustment

Model	Level α	Trend γ	Within-day seasonality δ	Within-week seasonality ω
Holt-Winters for Within-Day Seasonality	0.99	0.85	1	
Holt-Winters for Within-Week Seasonality	0.83	0		1
Double Seasonal Holt-Winters	0.85	0	0.53	1

Without residual auto-correlation adjustment

Table 2: Results Holt-Winters parameters wih autocorrelation adjustment

Model	Level α	Trend γ	Within-day seasonality δ	Within-week seasonality ω	AR ϕ
Holt-Winters for Within-Day Seasonality	0.80	0	0.69		0.74
Holt-Winters for Within-Week Seasonality	0.01	0		0.41	0.92
Double Seasonal Holt-Winters	0.01	0	0.18	0.31	0.94

With residual auto-correlation adjustment

It can be observed that the smoothing parameter for the level is estimated for the Holt-Winters models without residual auto-correlation adjustment are relatively close to 1. This indicates that the level of the current observation is given much more weight compared to the level of past observations, which implies that the current level estimate is highly being determined by the last observation. For the Holt-Winters method with residual auto-correlation adjustment it seems that the effect of the level is to a large extend being replaced by the auto-correlation adjustment. The smoothing parameter of the trend is estimated 0 for all methods, except for the Holt-Winters method for Within-Day Seasonality. This is probably due to the method's inability to capture the weekly pattern, which is has biased the trend estimation. This indicates that in general the trend effect is nihil. For the standard

Holt-Winters method, high values for the smoothing parameters for the seasonal indices can be observed, which is corresponding in line with my expectations that the time series is strongly dominated by the seasonal patterns. Again it can be observed that the estimated seasonal effects are smaller with the introduction of the residual auto-correlation adjustment. The results in this paper are in my opinion close to the results obtained by Taylor and in line with my initial expectations. Small differences are found for some parameters which can be due to some minor differences in the setup of the Holt-Winters method, such as initialization.

The estimated coefficients of the multiplicative ARIMA model are given by the following equation:

$$(1 + 0.98L^2)(1 + 0.78L^2)(1 + L^{336})y_t = -0.59 + (1 - 0.90L^2)z_t \tag{18}$$

The model contains an auto-regressive term for the second lag, a seasonal auto-regressive term for the second lag, a moving-average term for the second lag while the series is seasonaly differenced at seasonality equal to 336. The seasonality is captured by seasonally differencing the series. It can be observed from the resulting model that the electricity displays a strong hourly dependence of previous observations. The model was estimated for half-hourly lags as well, but this did not improve the model. It is not very clear why this is the case, but a possible reason might be that the electricity demand is recorded at the National Grid with hourly dependence. This model performs well in terms of mean-absolute percentage errors and serves more as a benchmark model to compare the other methods with.

The estimated coefficients of the full Fourier model are impractical to report, because of the large number of estimated paramaters involved in the model. The estimated parameters for the two partial Fourier models are also impractical for the same reason and interpretation is difficult. The following figure illustrates the estimation results by showing the forecasts of the first 2 weeks of the validation sample for both the full Fourier model and the Fourier model with a reduced number of harmonics, without the additional features of the model and the real data.

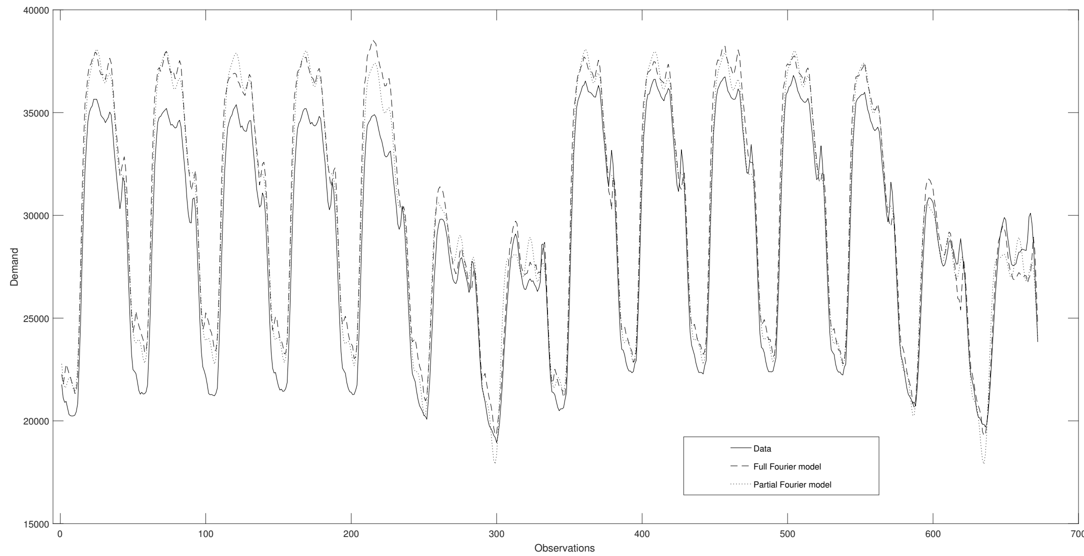

Figure 4: Forecast results of Fourier models

It can be observed that both models seem to capture the seasonal pattern of the data, because the model for the seasonality shows the same shape as the original data. The MAPE of the full Fourier model is given by 3.6% and the MAPE of the partial Fourier model is given by 3.29% for these 2 weeks of observations and forecasts for the validation sample. A remarkable observation is given by the fact that by modelling only the seasonal pattern of the data, the partial Fourier model provides better forecasts than the full Fourier model. There is no satisfactory explanation for this observed phenomenon in my opinion, since the partial Fourier model is a subset model of the full Fourier model. The full Fourier model is clearly over-fitting the data, but in my opinion the relative performance of the two models can not only be due to that reason. The most important thing that can be observed is the fact that both the full Fourier model an the Fourier terms of the subset Fourier model both seem to capture the seasonal pattern well.

4.2 Post-sample forecasting performance

The post-sample performance of the methods is compared by the mean absolute percentage errors (MAPE). Figure (5) shows the MAPE's for the Holt-Winters methods without auto-correlation adjustments and figure (6) shows the MAPE's of the Holt-Winters methods with autocorrelation adjustments. Both figures contain the MAPE's that were obtained from the forecasts of the SARIMA model.

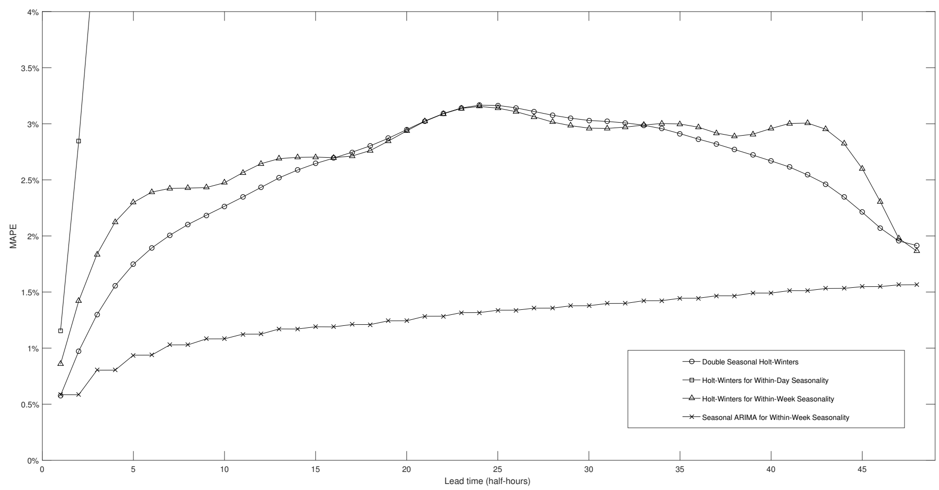

Figure 5: MAPES of models - no AR term

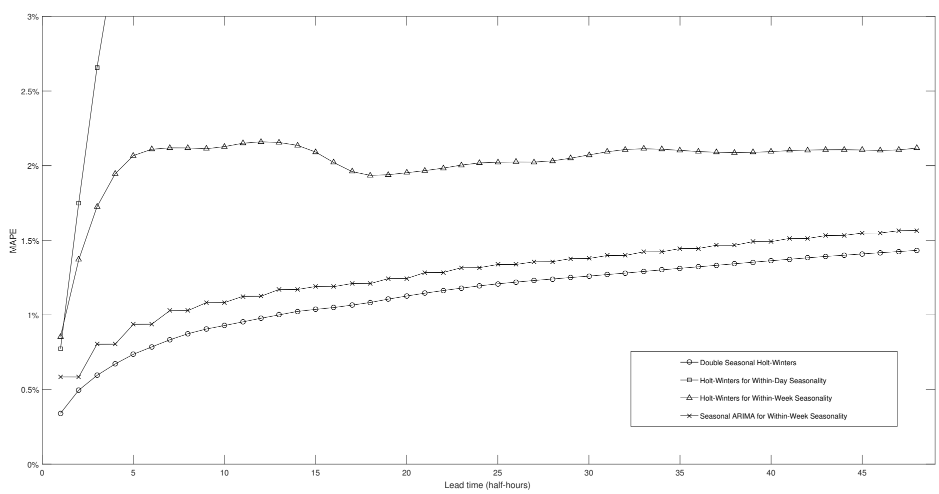

Figure 6: MAPES of models - AR term

It can be seen that the multiplicative seasonal ARIMA model clearly outperforms all Holt-Winters methods, including the double seasonal Holt-Winters method. The Holt-Winters method for within-day seasonality produces very inaccurate forecasts, to the extend that it could not be plotted in the figures. The double seasonal Holt-Winters method for within-day seasonality performs slightly better for low k and high k compared to the standard Holt-Winters method for withing-week seasonality. The models where the forecasts are adjusted, perform substantively better than models without auto-correlation correction of the residuals. Also it can be observed that the double-seasonal Holt-Winters method outperforms all other methods, including the well-specified multiplicative ARIMA model. These conclusions are similar to the conclusions made by Taylor (2003), but the MAPE's are somewhat different at some points. The models were run for the parameters obtained by Taylor, but this gave similar MAPE's as the ones displayed in the figures above. It is not clear to me where the differences in parameters estimates come from, but the small differences in the results might be due to differences in initialization.

The forecasting performance of the Fourier models is also evaluated by the MAPE's, which are presented in the next figure. The following figure displays the MAPE's of all Fourier models.

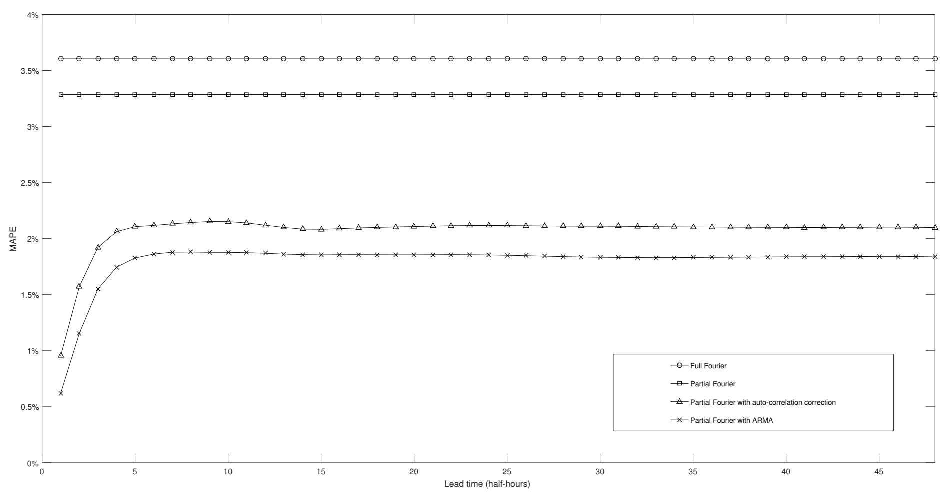

Figure 7: MAPE's of Fourier models

The first thing that can be observed from figure 7 is that the MAPE's of the full Fourier model are given by a horizontal line. This result is exactly as I expected, since forecasts are obtained by extrapolating the deterministic function of sine and cosine waves, which does not take into account information of new observations. The MAPE's of the two partial Fourier models both seem to converge to a certain number for increasing k when considering k-step ahead forecasts. The MAPE's of the error-correction model converge to a MAPE of 2.11% and the MAPE's of the combination of the partial Fourier model and the ARMA model converge to a MAPE of 1.83%. The convergence of the MAPE's can be explained by the way the model is setup. The MAPE's of the deterministic part of the models, which consists of trigonometric functions independent of new observations, are given by a constant number, as observed for the full Fourier model. The improvement of the constant MAPE is due to the inclusion of lagged observations or error terms in the model. These terms highly improve the short-term forecasts, but the effect decays for longer-term forecasts. This can be clarified by the following simple illustration. Consider a stochastic variable, say x_t which depends on the previous observation x_{t-1} with a certain effect λ between 0 and 1, i.e. $x_t = \lambda x_{t-1} + e_t$, then x_{t+k} can be written as follows:

$$x_{t+k} = \lambda^k x_t + e_{t+k} \qquad \text{for } 0 \leq \lambda \leq 1 \tag{19}$$

It can be seen from equation (19) that the effect of λ decays for increasing k because λ^k converges to zero. This effect is observed for the MAPE's of the partial Fourier models, because the effect of short-term lags decay.
It can be seen from figure 7 that the naïve full Fourier model can substantially be improved by adapting the model. The partial Fourier model with error-correction shows a MAPE that converges to 2.11% and the combined Fourier model performs slightly better with a MAPE that converges to 1.83%, which in my opinion is a good and promising result.

The following figure contains an overview of the MAPE's of the best performing models.

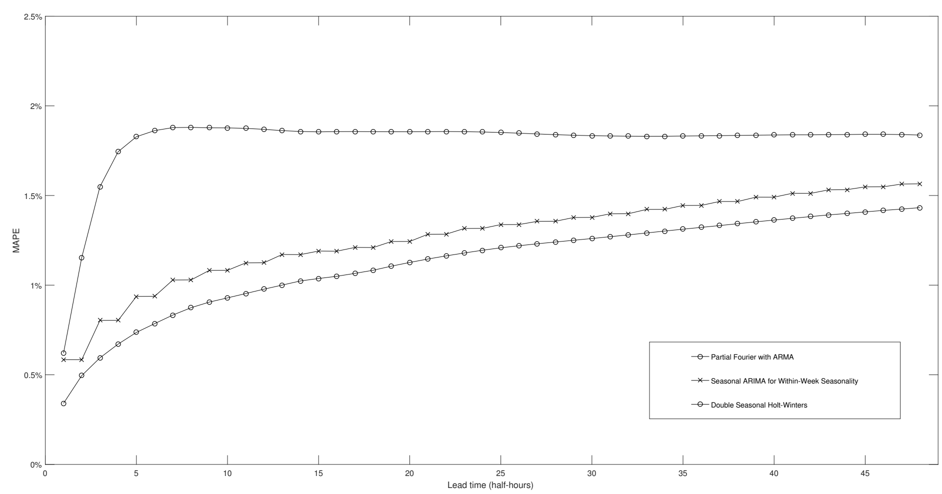

Figure 8: MAPE's of best models

Comparing all models together, I find the double seasonal Holt-Winters with residual auto-correlation correction the best model, in the sense that the method produces the most accurate forecasts of the validation sample. The multiplicative seasonal ARIMA is a very good competitor which performs well in practise if well-specified. The advantage of the Holt-Winters methods is that they are easy to implement and that they require little model specification efforts. The full Fourier model seems more of a naive method of forecasting that can substantially be improved by adapting the model. Correcting for residual auto-correlation yields improvement and the combination of the Fouriers series and an ARMA model performs almost as good as the double seasonal Holt-Winters method and the ARIMA model and is therefore a good alternative for the more widely-used models, since it has some nice features and is easy to work with.

5 Conclusion

Electricity demand forecasting is of great importance to balance supply and demand between generators and suppliers. Accurate electricity demand forecasts can lead to lower costs and more efficiency. The time series modelled in this paper considers a time series of half-hourly observations of electricity demand in England and Wales over a period of 12 weeks. The series contains strong seasonal patterns, therefore different forecasting methods from the time domain and the frequency domain are applied and compared with each other, where special attention is given to the seasonal patterns. A strong within-day seasonal cycle and a strong within-week seasonal cycle are evident from the data and these are analyzed and modelled from both the time-domain and the frequency domain perspective. I reach the same conclusion regarding exponential smoothing methods as Taylor (2003) who finds the double seasonal Holt-Winters method with residual auto-correlation adjustment the most accurate method, compared to other Holt-Winters methods and a well-specified multiplicative seasonal ARIMA model. Other methods have been developed and implemented by me that focus on modelling the electricity demand by trigonometric series. The models are given by the full Fourier model, a subset Fourier model with error-correction for the residuals and the most promising method where the series is modelled by a combination of trigonometric functions together with an ARMA proces. None of the methods were able to outperform the double seasonal Holt-Winters method with error-correction for the residuals, but last method performs almost as well as the ARIMA model and the double seasonal Holt-Winters method with residual auto-correlation correction. The methods discussed in this paper models a given time series by a deterministic function for seasonality together with stochastic components to account for short-term and long-term variation. The first advantage of the new model formulations is given by the fact that it can easily incorporate multiple non-integer seasonalities of different complexities. Another advantage is given by the fact that it can easily be extended with exogenous regressors and has a good adaptability. I conclude that the model has potential for modelling time series with evident seasonal patterns and I recommend research on the application of the model on medium-term and long-term time-series. I think think that the Fourier model formulations presented in this paper have strong potential for modelling time series with complex seasonalities, rather than modelling time series with simple seasonalities that can be captured with simpler models. If the trigonometric models presented in this paper proof to produce good results for other data sets, the model can be extended by adapting the model by methods such as the Box-Cox transformation, Furthermore I recommend that the relationship between the harmonics and the seasonal lags is investigated further and that a more formal procedure is specified for identifying the models. There will be situations where the model will perform very well and situations where the model is outperformed by other methods, therefore I recommended to always try to fit different models to the data.

References

[1] Bloomfield, P. (2004). Fourier Analysis of Time Series: an Introduction. John Wiley & Sons.

[2] Box, G. E., & Jenkins, G. M. (1976). Time Series Analysis, Control, and Forecasting. San Francisco, CA: Holden Day, 3226(3228), 10.

[3] Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.

[4] Chan, K. H. (1974). Business Time Series Spectral Forecasting Techniques. Xerox Microfilms.

[5] Chan, H., & Hayya, J. (1976). Spectral Analysis in Business Forecasting. Decision Sciences, 7(1), 137-151.

[6] Chatfield, C. (1978). The Holt-Winters Forecasting Procedure. Applied Statistics, 264-279.

[7] De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting Time Series with Complex Seasonal Patterns Using Exponential Smoothing. Journal of the American Statistical Association, 106(496), 1513-1527.

[8] Cleveland, R. B., Cleveland, W. S., & Terpenning, I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess. Journal of Official Statistics, 6(1), 3.

[9] Fourier, J. (1808). Mémoire sur la Propagation de la Chaleur Dans les Corps Solides. Nouveau Bulletin des Sciences de la Société Philomathique de Paris, 6, 112-116.

[10] Gardner, E. S. (1985). Exponential smoothing: The State of the Art. Journal of Forecasting, 4(1), 1-28.

[11] Gardner, E. S. (2006). Exponential smoothing: The State of the Art—Part II. International journal of forecasting, 22(4), 637-666.

[12] Gilchrist, W. (1976). Statistical Forecasting, London Wiley-Interscience Pub.

[13] González-Romera, E., Jaramillo-Morán, M. A., & Carmona-Fernández, D. (2008). Monthly Electric Energy Demand Forecasting with Neural Networks and Fourier Series. Energy Conversion and Management, 49(11), 3135-3142.

[14] Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). Forecasting with Exponential Smoothing: the State Space Approach. Springer Science & Business Media.

[15] Koopmans, L. H. (1995). The Spectral Analysis of Time Series. Academic press. ISO 690

[16] Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., ... & Winkler, R. (1982). The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition. Journal of forecasting, 1(2), 111-153.

[17] Makridakis, S., Hibon, M., & Moser, C. (1979). Accuracy of Forecasting: An Empirical Envestigation. Journal of the Royal Statistical Society. Series A (General), 97-145.

[18] McLoughlin, F., Duffy, A., & Conlon, M. (2013). Evaluation of Time Series Techniques to Characterise Domestic Electricity Demand. Energy, 50, 120-130.

[19] Rayner, J. N. (1971). Introduction to Spectral Analysis. Pion.

[20] Reid, D. J. (1975). A Review of Short-Term Projection Techniques. Practical Aspects of Forecasting, 8-25.

[21] Taylor, J. W. (2003). Short-Term Electricity Demand Forecasting Using Double Seasonal Exponential Smoothing. Journal of the Operational Research Society, 54(8), 799-805.

[22] Taylor, J. W., & Buizza, R. (2003). Using weather ensemble predictions in electricity demand forecasting. International Journal of Forecasting, 19(1), 57-70.

[23] Taylor, J. W., De Menezes, L. M., & McSharry, P. E. (2006). A Comparison of Univariate Methods for Forecasting Electricity Demand up to a Day Ahead. International Journal of Forecasting, 22(1), 1-16.

[24] Williams, D. W., & Miller, D. (1999). Level-Adjusted Exponential Smoothing for Modeling Planned Discontinuities. International Journal of Forecasting, 15(3), 273-289. ISO 690