3rd International Conference on Industrial Engineering and Industrial Management
XIII Congreso de Ingeniería de Organización
Barcelona-Terrassa, September 2nd-4th 2009

Load demand forecasting from one day to several months ahead based on State Space models. An empirical evaluation

Juan R. Trapero Arenas[1]

[1] Dpt. Management Science. Management School. Lancaster University. LA1 4YX , Lancaster (UK). j.traperoarenas@lancaster.ac.uk.

Keywords: load demand forecasting, Unobserved Component Models, Fixed Interval Smoothing, multi-rate forecasting.

1. Introduction

Deregulation processes during the last two decades across many developed economies have motivated the need for more accurate forecasting tools of electricity markets. Load demand prediction is important for the development of any model for electric power system planning. Medium term forecasts (one day to several months) are typically used to schedule fuel purchases reducing financial risks and for maintenance operations, Yalcinoz and Eminoglu (2005) *.

Most of the literature on load demand forecasting refers to short-term forecasting, up to one day or one week ahead at most. In that context it is generally acknowledged the inherent modelling difficulties, like high frequency observation load data (usually hourly, but often half-hourly); the presence of multiplicity of periodic behaviour superimposed (strong daily and weekly patterns); atypical effects (holidays, public holidays, television events, etc.); non-linear relations with other variables, mainly weather variables; etc. This complexity is reflected in the high number of methods applied to the problem, ranging from classical methods (like regression, Exponential Smoothing, ARIMA, optimal Kalman filtering, etc.) to some considered more modern, like Artificial Neural Networks, which have become almost the rule, overshadowing other possibilities. Some references are Taylor et al. (2006) and Pedregal and Trapero (2007).

It is obvious that the difficulties in forecasting load demand for long (more than a year ahead) or mid-term (one day to several months) are much more important than for short-term, and this is reflected in a much scarcer amount of literture, Amjady and Keynia (2008).

The references on long or mid-term forecasting are very heterogeneous in many respects, but two are most important from the point of view of this paper. Firstly, the spread of methods implemented is considerable, as it is expected when the issue addressed is rather complex. Secondly, and most importantly, the nature of the data used, mainly their sampling interval, is very different from some references to others.

The aim of this paper is developing a general multi-rate methodology in order to find forecasts as accurate as possible for a mid-term horizon (up to twelve weeks in the examples below, about three months) for data sampled at an hourly rate. This hourly

* This work has been partly supported by La Consejería de Educación y Ciencia de la Junta de Comunidades de Castillla-La Mancha under Project PCI08-0042-6312 and by a Marie Curie Intra European Fellowship within the 7th European Community Framework Programme (FP7-PEOPLE-IEF-2008)

sampling interval then restricts considerably the number of similar applications found in the literature. The forecasting horizon may be extended as long as it is sensible to do so, depending on the amount and quality of the data, but it is not a constraint imposed by the method. In addition, this work extends the short-term forecasting technique presented in Pedregal and Trapero (2007) for a mid/long term version applied to load demand data.

The outline of the paper is as follows: section 2 presents the general UC framework in which our models are set up and discusses the particular model for any of the components involved in the case of load demand; section 3 presents the particular multi-rate approach proposed; section 4 exposes a collection of empirical findings; and section 5 sums up and extracts the most important conclusions.

2. Unobserved Components models

Unobserved Component models (UC) are a class of stochastic processes that has proven very useful in a wide range of scientific areas, Young et al. (1999). There are some specific applications to forecasting electricity markets, Pedregal and Trapero (2007); but in general publications are scarce, especially when compared to other approaches.

In a univariate UC model, the signals are assumed to be the addition of several components, each one with its own physical interpretation. Since the components are not directly observable, there are many ways of decomposing the time series. To remove this ambiguity, assumptions have to be made about the statistical nature of the components. In the case of a monthly time series with a seasonal component, a typical model is

$$y_t = T_t + C_t + S_t + e_t \quad (1i)$$

where y_t is the observed time series; T_t is a trend or low frequency component; C_t is a possible cyclical component with a period longer than one year; S_t is the seasonal component with a period of 12 observations/months per year; and e_t is a serially uncorrelated white noise (with constant variance σ^2).

The model ought to be more complicated for hourly electricity demand; one fairly general formulation is

$$y_t = T_t + D_t + W_t + A_t + e_t \quad (1ii)$$

where t is now measured in hours; D_t is a daily periodic component; W_t is a weekly component; and A_t is an annual component. The annual cycle or seasonality A_t is of paramount importance in the present context, when forecasts for mid-term at an hourly interval are required. When the forecasting horizon is no longer than one week, the term A_t is usually dropped from any model, but for longer horizons this term is essential to get sensible forecasts, as it will be shown later on.

A different way of writing equations (1) is given in equation (2), where $C_t + S_t = \sum_{k=1}^{P/2} S_{k,t}$ for (1i) and $D_t + W_t + A_t = \sum_{k=1}^{P/2} S_{k,t}$ for (1ii).

$$y_t = T_t + \sum_{k=1}^{P/2} S_{k,t} + e_t \quad (2)$$

The only difference between both options is that different values for P are necessary in order to reach an appropriate representation of the time series, where P is the period of the longer periodical component measured in the sampling rate, i.e. $P = 12$ months for (1i) and $P = 8{,}760$ hours in (1ii). It is obvious that in the latter case the number of sub-components in the sum (2) is immense, exactly 4,380, and the estimation problem simply blows up because of the model dimension and the number of parameters involved. Therefore, component A_t has to be necessarily removed from the model due to technical problems, but still it is essential for obtaining sensible mid-term forecasts. The solution to this problem is the topic of this paper and is presented and tested in the following sections. When A_t is dropped off the equation (1ii) or (2) the longer period is the weekly one, that is $P = 168$ hours, still high, but manageable with the estimation procedures in the frequency domain proposed below.

At this point in time, UC and State Space (SS) models may be considered classical techniques, and therefore the main topics about state and parameter estimation will be briefly summarised here. Readers with more interest on these topics may consult Harvey (1989).

In a standard State Space framework, equation (1) or (2) is considered as the observation equation which describes the stochastic evolution of state variables associated with the UC's in (1). The SS description of the full UC model is obtained by assembling all the individual SS forms of all the components. Therefore, in order to formulate this overall SS form of the model, specific assumptions about the statistical nature of every component have to be made. The adequacy of these assumptions may be checked afterwards, by standard testing procedures.

All the components in the models shown so far are basically trends and periodical components of different frequencies/periods. The SS representation of each of them used in this paper is the typical of the so called Basic Structural Model (BSM), Harvey (1989), and is given in equations (3) and (4).

$$\begin{pmatrix} T_t \\ T_t^* \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} T_{t-1} \\ T_{t-1}^* \end{pmatrix} + \begin{pmatrix} w_t \\ w_t^* \end{pmatrix} \quad (3)$$

$$\begin{pmatrix} S_{k,t} \\ S_{k,t}^* \end{pmatrix} = \begin{pmatrix} \cos\frac{2\pi k}{P} & \sin\frac{2\pi k}{P} \\ -\sin\frac{2\pi k}{P} & \cos\frac{2\pi k}{P} \end{pmatrix} \begin{pmatrix} S_{k,t-1} \\ S_{k,t-1}^* \end{pmatrix} + \begin{pmatrix} w_{k,t} \\ w_{k,t}^* \end{pmatrix} \quad (4)$$

$$k = 1,2,\ldots P/2$$

Here the trend noise w_t, w_t^* are random Gaussian noises, independent of each other with zero mean and certain variances σ_T^2 and σ_T^{2*}, respectively; and the noises $w_{k,t}$ and $w_{k,t}^*$ are independent random noises with common variance σ_k^2 (but $\sigma_k^2 \neq \sigma_j^2$ for any $k \neq j$ with $j,k = 1,2,\ldots,P/2$). Another typical representation is the so called Dynamic Harmonic Regression (Young et al. (1999)).

By selecting certain sum of subsets of the terms $S_{k,t}$, meaningful components may be defined. A daily cycle for the hourly sampling rate data may be estimated by adding up the seasonal sub-components corresponding to the daily frequency and its harmonics, i.e.

$D_t = \sum_{k=1}^{12} S_{k,t}$ for $k = 7j$ and $j = 1,2,\ldots,12$. In a similar way, a weekly cycle may be found by summing up the weekly periodic term and all the harmonics not included in the daily cycle.

Given the full SS system, formed by assembling equations (2), (3) and (4) it is well known that the Kalman Filter (KF) and the Fixed Interval Smoothing algorithms (FIS) provide the optimal estimation of the first and second order moments of the state vector in the sense of minimizing the Mean Squared Error. They also produce automatically a number of important operations in time series analysis, like interpolation of missing observations, forecasting, backasting (if necessary), etc.

The application of the recursive KF and FIS algorithms requires the knowledge of all the system matrices, something that is not known in general. In the system above, the unknown parameters are all the noise variances, i.e. σ_T^2; σ_T^{2*}; σ_k^2 with $k = 1,2,\ldots,P/2$; and σ^2, the irregular/innovations variance. In many applications of single output UC systems the number of unknown parameters may be reduced by one, by normalizing all the variances by the innovations variance. In this way, Noise Variance Ratios (NVR) are usually defined as σ_T^2/σ^2; σ_T^{2*}/σ^2; σ_k^2/σ^2 with $k = 1,2,\ldots,P/2$. When this change is done, the KF and FIS algorithms have to be updated accordingly.

The estimation method usually preferred in the literature, due to its general good statistical properties is Maximum Likelihood (ML). However, it is well known that the likelihood surface is very flat or multimodal around the optimum when a big number of parameters have to be estimated, as it is reflected on the fact that the searching algorithms never converge to a clear optimum; they converge to different optima, depending on the initial conditions; or the standard errors of estimates are usual extremely high.

The likelihood function in the frequency domain may be obtained by a formal translation of ML in time domain based on a Fourier transform which converts serial correlation into heteroscedasticity (Harvey, 1989). Fortunately, for time series with a marked periodic behaviour, the likelihood function defined in the frequency domain is much better defined than in time domain, even for high dimensional models. This will be the estimation method used later on.

3. Multi-rate approach

One important limitation of the previous models for data sampled at an hourly rate when mid or long-term forecasts are required is that they do not incorporate the obvious annual seasonal pattern (i.e. A_t in equation (1ii)). It is generally acknowledged in many publications that avoiding this fact is unimportant for short-term forecasting (up to one week ahead). On the contrary, it is essential for longer forecasting horizons, and therefore it should be incorporated necessarily.

There are many options to fit in that component into the model, but one option that is feasible and efficient is to forecast the time series at different sampling intervals and link both sort of forecasts by means of time aggregation techniques. The procedure consists of two broad steps that will be explained later in a more detailed algorithm:

1. Find optimal forecasts for the next required months on the basis of monthly data and models that incorporate the annual seasonality explicitly.
2. Build a model for the hourly data typical of short-term forecasting horizons and forecast the hourly time series for the next months with this model, but making sure that certain constraints are fulfilled. Such constraints are that the sum of the

forecasted hourly values for each month are exactly equal to the monthly forecasts found in the previous step.

The main technical problem here is finding the forecasts in step 2. with the required constraints. However, this is relatively straightforward if the model in step 2. may be written in State Space form by means of time aggregation techniques, that in the SS framework are particularly natural, as described below. Indeed, provided an appropriate SS representation of the hourly series is found, the KF and FIS algorithms provide the forecast required.

Let's assume that the BSM model set up in the SS form given by equations (2), (3) and (4) is written in compact form as equation (5), where $\mathbf{x}_t$ is the general state vector, i.e. the trend, the periodic components and all the auxiliary states; $\mathbf{\Phi}$, $\mathbf{E}$ and $\mathbf{H}$ are the system matrices, formed by block concatenation of the individual system matrices given in (2), (3) and (4). The appendix shows the full BSM model used in later examples according to this notation.

$$\begin{cases} \mathbf{x}_t = \mathbf{\Phi}\mathbf{x}_{t-1} + \mathbf{E}\mathbf{w}_t \\ y_t = \mathbf{H}\mathbf{x}_t + v_t \end{cases} \tag{5}$$

In order to set up model (5) in which the temporal aggregation is taken into account an explicit cumulator variable has to be defined, see Harvey (1989)). A first step is setting up the previous model including the observation equation into the state vector, i.e.

$$\begin{cases} \begin{bmatrix} y_t \\ \mathbf{x}_t \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{H\Phi} \\ \mathbf{0} & \mathbf{\Phi} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \mathbf{x}_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & \mathbf{HE} \\ \mathbf{0} & \mathbf{E} \end{bmatrix} \begin{bmatrix} v_t \\ \mathbf{w}_t \end{bmatrix} \\ y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} y_t \\ \mathbf{x}_t \end{bmatrix} \end{cases} \tag{6}$$

Beware that system (6) is exactly equivalent to (5). The second step is incorporating the cumulator variable to this model that will produce the required constraints. This is a variable defined as

$$C_t = \begin{cases} 0, & \begin{array}{c} t = \text{every hour within the estimation sample and} \\ \text{first hour of each month into the prediction sample.} \end{array} \\ 1, & \text{otherwise.} \end{cases}$$

The final model is then (7).

$$\begin{cases} \begin{bmatrix} y_t \\ \mathbf{x}_t \end{bmatrix} = \begin{bmatrix} C_t & \mathbf{H\Phi} \\ \mathbf{0} & \mathbf{\Phi} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \mathbf{x}_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & \mathbf{HE} \\ \mathbf{0} & \mathbf{E} \end{bmatrix} \begin{bmatrix} v_t \\ \mathbf{w}_t \end{bmatrix} \\ y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} y_t \\ \mathbf{x}_t \end{bmatrix} \end{cases} \tag{7}$$

Model (7) has some peculiarities, (i) there is no observed noise, (ii) the first state replicates the observed data exactly; (iii) while system (6) is time invariant, in (7) there is one time varying system matrix due to the introduction of the cumulator variable; (iv) the first state of the system is the data for the estimation sample, but is an accumulated version of the output in the forecasting sample.

The way the forecasts are produced in the State Space framework defined may be summarised in the following algorithm:

1. Find optimal monthly forecasts based on monthly observations and an optimal procedure. In later examples a monthly BSM will be used.
2. Set up and estimate the unknown parameters in the frequency domain on a BSM model for hourly data with a weekly cycle, i.e. model (1ii) without the annual cycle A_t (see the appendix).
3. Form a new time series by appending two sub-series. The first part is simply the actual hourly data in the estimation sample. The second part is the monthly forecasts from model in step 1. allocated in the appropriate hourly sampling scale, with missing observations in the middle.
4. Run the Kalman Filter and Fixed Interval Smoothing on the time series formed in step 3. with system (7) and the parameter estimates from step 2.
5. Since the first state is the accumulated output of the system in the forecast period, the final forecasts ought to be built by performing the reverse operation to accumulation done by the SS model. This amounts to make a first difference of the first state in the forecasting sample, for most hours, with the exception of those hours where $C_t = 0$.

The role of the cumulator variable C_t in system (7) is the key point in this procedure. Since the first state of the system is the output (i.e. the data), whenever this variable takes a value of one the first state is accumulating the values of the output measured at that hour and all the previous hours. Such accumulation is broken or re-started as soon as its value becomes zero. Then, given the distribution of zeros for C_t (the first hour of each month in the forecast period) the Fixed Interval Smoothing algorithm will produce a forecast value for the end of each month that is exactly the value of the output the algorithm finds at that point in time. But those values have been previously set as the values of the monthly forecasts from the monthly model. In this way the monthly constraints are preserved.

It is important to note that the procedure is very general, thanks to the State Space formulation. The only restriction is that the model for the data in the shorter sampling interval should be written in State Space form, the rest of the method applies automatically. This means that it could be applied to any other combination of sampling intervals and that models in either timing may be of any kind preferred by each analyst. In particular, models in the coarser sampling interval may be non linear, incorporate inputs, add on any kind of judgment, etc. While the model in the finer sampling interval could be either of the kind used in this paper or ARIMA, Exponential Smoothing, etc.

There are at least two factors that make the problem especially difficult in technical terms, as it is presented in this paper. Firstly, the dimension of the hourly BSM model, since we need explicitly the estimation of much more parameters than in standard applications of this model. This fact motivated the estimation by ML in the frequency domain. Secondly, the specific properties of system (7), mainly that one system matrix is time varying. The result of both facts is that there is no commercial software available in the market to solve the problem and this motivated the development of our own software, written in MATLAB™.

4. Empirical results

The data used to illustrate the method proposed in the previous sections are the recent hourly electricity load data registered on a transformer of an important electrical company in the UK (37,753 observations), see Figure 1.

Two are the main issues addressed in this paper. Firstly, how big is the hourly forecast improvement by using a model that includes the annual cycle with respect to standard alternatives typical of short-term forecast applications. Secondly, we look for the forecast horizon from which the improvements start to be important.

In order to find the evidence, a forecasting experiment was set up. The experiment consists of the application of the algorithm in a rolling manner along a full year of data. The forecast horizon was fixed to 2,016 hours ahead (12 weeks, about three months), but longer forecasting horizons could be used. Once a forecast is done, the forecast origin is moved one day ahead, until completing 365 sets of three months ahead forecasts. The estimation of the models is updated every day. Additionally and for forecasting comparison purposes, a model set up for hourly data typical of short term forecasting (without the annual cycle) was estimated in the same way as the proposed algorithm.

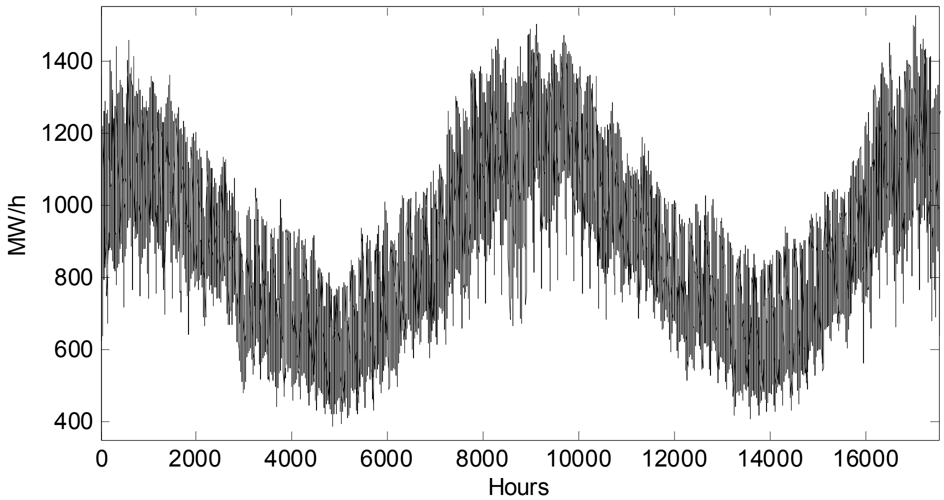

Figure 1. Hourly load demand for two years registered at a transformer of a UK company

The prediction error measure chosen for this work was the well known Mean Absolute Percentage Error (MAPE), given in equation (8), where $\hat{z}_{t+i}$ stands for the forecasting values at time $t+i$; z_{t+i} are the actual load demand values; and n takes values from 1 to 2,016 hours.

$$MAPE(n) = \frac{1}{n} \frac{\sum_{i=1}^{n} \left| \hat{z}_{t+i} - z_{t+i} \right|}{z_{t+i}} \cdot 100 \tag{8}$$

Figure 2 presents the forecasting performance of the algorithm (UC1 from now on), compared to the simpler alternative, typical of short term contexts (UC). In order to carry out this comparison, robust statistical measures to atypical observations like the median and mad (i.e. the median absolute deviation with respect to the median) were used. Each line in the figure represents the median of the MAPE for the whole set of 365 forecast errors from 1 to 2,016 hours ahead obtained along the year. Vertical dotted lines indicate the beginning of each week. In a similar format, Figure 3 also shows the mad.

Several conclusions may be extracted from all this information. Firstly, there is a rapid increase in the error measurements in all models for very short forecasting horizons, but the increments are reduced considerable after half a week for the UC1 model, while UC still grows at a high rate. Secondly, it is verified that the inclusion of the annual cycle in the model is very important, because the error is reduced to almost a half for a three month forecast regarding the median and more than a half, looking at the mad. In other words, forecasts are both more accurate and their dispersion is much lower. Thirdly, the improvement consistently increases with the forecasting horizon. Finally, including the annual cycle starts to produce consistent improvements for horizons of one week or longer. Certainly, for horizons up to one day ahead forecasts, typical of short term applications the inclusion of the annual cycle does not produce any problem, but is not worthy in forecasting terms

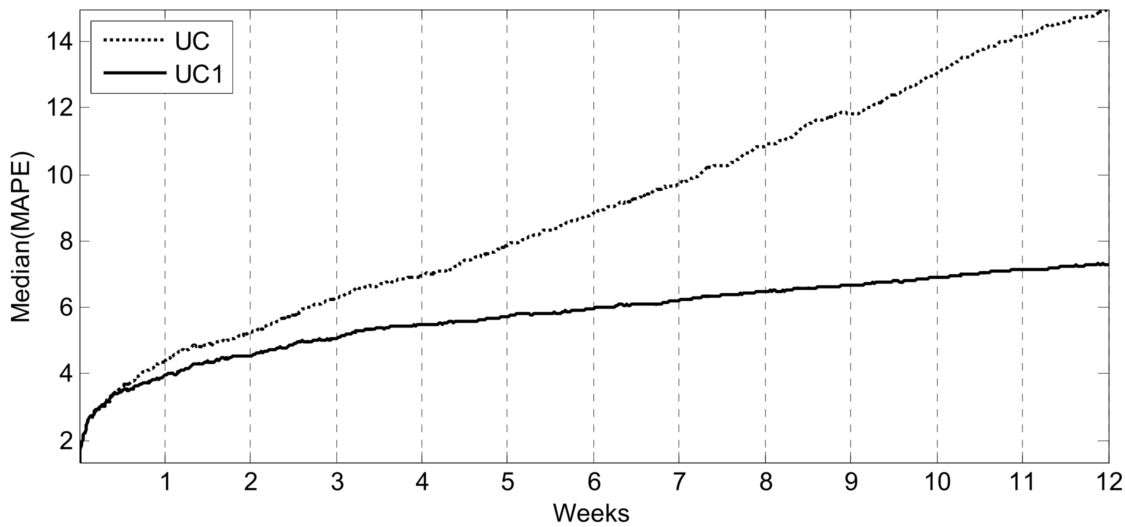

Figure 2. Median of MAPE of the 365 sets of 12 week ahead forecast errors for the two models considered.

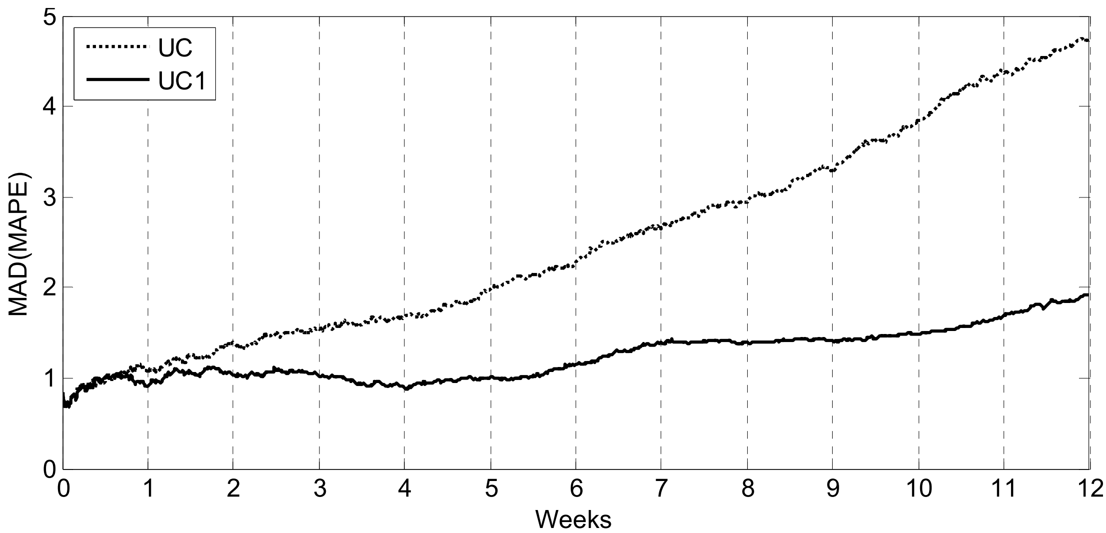

Figure 3. MAD of MAPE of the 365 sets of 12 week ahead forecast errors for the two models considered.

5. Conclusions

This paper proposes a multi-rate general method for long-term forecasting electricity load demand for rapidly sampled data. The rapid sampling interval of the original data and the forecasting horizon make the problem especially difficult, since the obvious annual seasonality cannot be avoided, as it is usually done in short-term forecasting studies (normally up to one day ahead).

The approach consists of a two step efficient procedure in which a monthly forecast is produced on the basis of monthly data and such forecasts are fixed as constraints for an hourly model estimated on hourly data. The monthly model incorporates the required annual seasonality by an appropriate model. The rest, i.e. the more detailed information typical of rapid sampled data, is produced by the hourly model. Imposing the constraints by the forecasts in the first step is natural to implement by means of a standard State Space framework. The method is very general in the sense that any type of model could be used in both steps, allowing the analysts to incorporate their own experience with particular time series or models. The only restriction for the method to work is that the hourly model should be written in State Space form. Unobserved Component Models have been used in this paper in both stages.

The procedure is evaluated by a thorough forecasting experiment in which 365 sets of one hour up to twelve weeks ahead of hourly forecasts are produced for the load demand registered at a transformer of a UK company. The method is compared with the model in the second step, in which the annual seasonality is not incorporated. The conclusions from the experiment are very clear:

- Modelling the annual seasonality reduces the forecast error to a half in the horizon of three months ahead, both measured as the median or the mad of the Mean Absolute Percentage Error.
- Forecasting advantages of the method start to be important for horizons of one week ahead and longer.

References

Yalcinoz, T.; Eminoglu, U. (2005). Short term and medium term power distribution load forecasting by neural networks. Energy Conversion & Management, Vol. 46, pp. 1393-1405.

Taylor, J.W.; 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, Vol. 22, pp. 1-16.

Pedregal, D.J.; Trapero, J.R. (2007). Electricity prices forecasting by automatic dynamic harmonic regression models. Energy Conversion & Management, 48, 1710-1719.

Amjady, N.; Keynia, F. (2008). Mid-term load forecasting of power systems by a new prediction method. Energy Conversion & Management, Vol. 49, pp. 2678-2687.

Young, P. C.; Pedregal, D. J.; Tych, W. (1999). Dynamic Harmonic Regression, Journal of Forecasting, Vol. 18, pp. 369-394.

Harvey, A. C. (1989), Forecasting Structural Time Series Models and the Kalman Filter. Cambridge: Cambridge University Press.