Short-Term Load Forecasting with Exponentially Weighted Methods

James W. Taylor

IEEE Transactions on Power Systems, 2012, Vol. 27, pp. 458-464.

***Abstract*—Short-term load forecasts are needed for the efficient management of power systems. Although weather-based modeling is common, univariate models can be useful when the lead time of interest is less than one day. A class of univariate methods that has performed well with intraday data is exponential smoothing. This paper considers five recently developed exponentially weighted methods that have not previously been used for load forecasting. These methods include several exponential smoothing formulations, as well as methods using discount weighted regression, cubic splines and singular value decomposition (SVD). In addition, this paper presents a new SVD-based exponential smoothing formulation. Using British and French half-hourly load data, these methods are compared for point forecasting up to one day ahead. Although the new SVD-based approach showed some potential, the best performing method was a previously developed exponential smoothing method. A second empirical study showed the better of the univariate methods outperforming a weather-based method up to about five hours ahead, with a combination of these methods producing the best results overall.**

***Index Terms*— Discount weighted regression, exponential smoothing, load forecasting, singular value decomposition, spline functions.**

I. Introduction

Automated short-term load prediction is needed for the efficient operation of power systems. It is also used to support transactions by participants in deregulated electricity markets [1],[2]. Although weather-based models are often used to predict load, they are less important for short horizons, as weather variables tend to change relatively smoothly over short intervals of time. This prompts consideration of modeling approaches that use only historical load data. These are termed univariate methods. A common approach to short-term load forecasting is to decompose the load into weather sensitive and weather insensitive components [3], and for the latter univariate methods are required. Of course when weather data and forecasts are not available, or are prohibitively expensive, a univariate method must be used.

The key feature of intraday load data that must be accommodated in a univariate method is the existence of intraday and intraweek seasonal cycles. Two univariate methods that have performed well in empirical studies are HWT exponential smoothing, which is the Holt-Winters method adapted for the modeling of two seasonal cycles, and an intraday cycle (IC) exponential smoothing method [4], [5]. The success of these methods, and the common use of exponential smoothing in business and industry, motivates the consideration in this paper of five exponentially weighted methods recently developed in [6] for intraday time series. In [6], the accuracies of the methods were compared using a 35-week call center time series for prediction up to two weeks ahead. In this paper, the main empirical analysis involves two load series, each consisting of three years of half-hourly observations, which are used to evaluate point forecast accuracy from one half-hour up to 24 hours ahead. An additional empirical study is presented that compares the univariate methods to a weather-based approach.

The first of the five methods from [6], considered in this paper, involves exponential smoothing of both the total load for a week and the partition of this weekly total across the periods of the week. The second method is discount weighted regression (DWR) with trigonometric terms. This is a development of an idea put forward in this journal 40 years ago [7]. That paper used exponentially weighted regression. The development provided by DWR is the use of more than one discount factor. The third method from [6] uses DWR to fit a time-varying regression spline to the seasonal cycles. The fourth method is an alternative form of time-varying spline, which uses exponential smoothing to model the spline at the knots. The fifth method aims to reduce the dimensionality of the modeling of intraday data by using singular value decomposition (SVD) to transform the data before the use of exponential smoothing. The availability of a long time series of intraday observations prompts the development in this paper of a new and simpler version of this method. Therefore, although the contribution in this paper is largely empirical, it also provides this methodological development.

This paper's empirical analysis includes, as benchmarks, the HWT and IC exponential smoothing methods, as well as an autoregressive integrated moving average (ARMA) model, and an artificial neural network (ANN). The short-term load forecasting literature contains a variety of ARMA models, including periodic [5] and double seasonal [8], [9]. Although the nonlinear and nonparametric features of ANNs are particularly attractive for weather-based load modeling, univariate ANNs have also been applied to load [e.g. 10].

Manuscript received February 9, 2011.

J. W. Taylor is with the Saïd Business School, University of Oxford, Park End Street, Oxford OX1 1HP, UK (e-mail: james.taylor@sbs.ox.ac.uk).

Section II introduces the data used in the main empirical study in this paper. Section III presents the exponentially weighted methods. Section IV describes the ARMA and ANN methods. Section V provides an empirical comparison of the univariate methods. Section VI presents an additional empirical study that includes a weather-based method. Section VII provides a summary and conclusion.

II. The Load Series

The main empirical analysis in this paper compares forecasting methods using a British and a French load series. Both consist of all half-hourly observations in 2007, 2008 and 2009. The methods were fitted using the first two years of data, and post-sample accuracy was evaluated for the final year. The forecast origin was rolled through 2009 to deliver a set of forecasts for one half-hour up to 24 hours ahead.

Fig. 1 shows a summer and winter fortnight from the British series. Each fortnight shows two reasonably similar intraweek cycles of length m_2=7×48=336 half-hourly periods. There is also similarity between the intraday cycles of length m_1=48 periods, particularly for the weekdays. These features were also evident in the French data, and are typical of intraday load data. The univariate methods in this paper aim to model these two types of seasonal cycle. The seasonality is noticeably different in the winter and summer fortnights. Therefore, methods are needed that model the seasonality as time-varying.

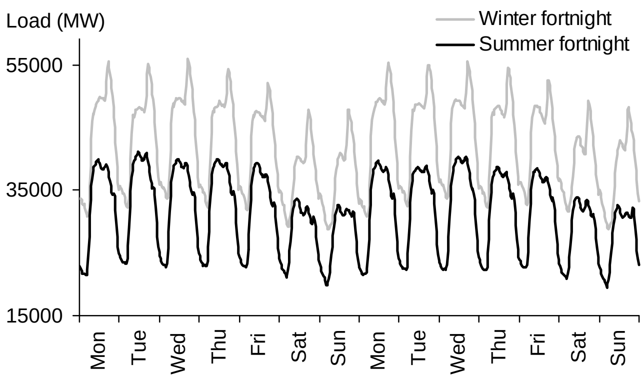

Fig. 1. British load for a winter fortnight (12 to 25 January 2009) and a summer fortnight (20 July to 2 August 2009).

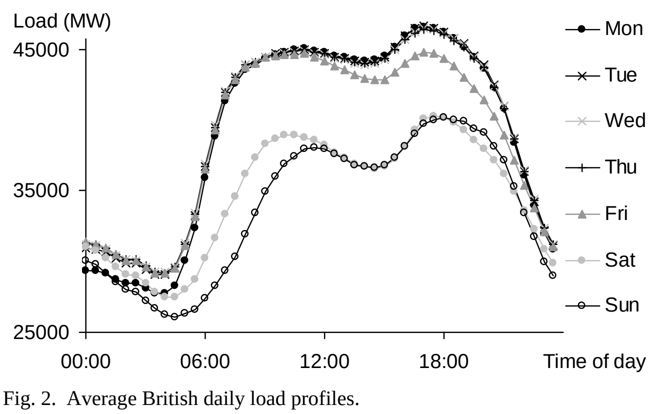

Fig. 2. Average British daily load profiles.

For each day of the week, Fig. 2 shows the British load profile averaged over data from our estimation sample. This figure emphasizes the differences and similarities between the patterns of load for different days of the week.

The two series both possess strong annual seasonality, with lower demand during the warmer months of each year. The exponentially weighted methods proposed in [6] could be developed to accommodate the annual cycle. For the exponential smoothing models, this could be achieved using the approach in [11], but other ideas would be needed for the methods based on DWR. In this paper, the annual seasonality is not incorporated in the methods. Modeling the annual cycle would clearly be important for forecasting at lead times of several years, months and weeks, but it is less important when interest is, as in this paper, in prediction up to just a day ahead.

Prior to modeling the series, the natural log transformation was applied in order to stabilize the variance of each series, and because the models to be considered have additive structure. Both series possessed days for which the pattern of load was unusual, such as public holidays. As the methods in this paper are not suitable for such days, observations for these days were smoothed out before modeling. For the in-sample data, this generally involved replacing an unusual observation by the average of the load in the corresponding periods of the two adjacent weeks. The unusual days were not included when optimizing parameters and evaluating post-sample forecasts.

III. Exponentially Weighted Methods

A. HWT Exponential Smoothing

The HWT method, introduced in [12], extends Holt-Winters exponential smoothing with the aim of modeling the intraday and intraweek cycles in intraday data. The method is presented here in full to assist later discussions. The method can be presented as the following state space model:

$$y_t = l_{t-1} + d_{t-m_1} + w_{t-m_2} + \phi e_{t-1} + \varepsilon_t \quad (1)$$

$$e_t = y_t - \left(l_{t-1} + d_{t-m_1} + w_{t-m_2}\right) \quad (2)$$

$$l_t = l_{t-1} + \alpha e_t \quad (3)$$

$$d_t = d_{t-m_1} + \delta e_t \quad (4)$$

$$w_t = w_{t-m_2} + \omega e_t \quad (5)$$

where $\varepsilon_t \sim \mathrm{N}(0,\sigma^2)$ and σ^2 is a constant variance. Throughout this paper, y_t is the log transform of load. l_t and d_t are the state variables for the level and intraday cycle, respectively; w_t is the state variable for the intraweek cycle remaining after d_t is removed; α, δ and ω are smoothing parameters; and the term involving parameter ϕ is an important adjustment for residual autocorrelation. The inclusion of a trend term was considered, but this did not improve accuracy. Monte Carlo simulation can be used to generate prediction intervals, and indeed point forecasts, from the model. For all exponential smoothing methods in this paper, the initialization of state variables and the least squares parameter optimization were performed using the procedures described in [6]. The optimized values for the British data were: α=0.001, δ=0.302, ω=0.399 and ϕ=0.969.

B. IC Exponential Smoothing

Intraday cycle (IC) exponential smoothing is presented in [4]. Although the IC method smoothes the level in the same way as the HWT method, these methods differ noticeably in how they model the seasonality. The IC method has no term for the intraweek cycle. It allows the same intraday cycle to be used for days of the week with similar load profiles. In view of Fig. 2, and the analogous plot for the French data, for both series, an IC model was implemented with distinct intraday cycles for Monday, Friday, Saturday and Sunday, and a common intraday cycle for the other days of the week. As in [5], the accuracy of the IC model substantially improved with the inclusion of a similar residual autocorrelation term to that included in the HWT model. The two forms of the IC model evaluated in [5] were both implemented in this paper. The 'restricted' form constrains some of the parameters to be identical, while the 'unrestricted' form only constrains the parameters to be nonnegative and less than 1.

C. Total and Split Exponential Smoothing

This method is introduced in [6]. It smoothes both the weekly total and the split of this total across the periods of the day and the week. The smoothing of the weekly total is an interesting alternative to the HWT and IC methods, which involve the more classical smoothing of a level term.

D. DWR with Trigonometric Terms

In a linear regression framework, a model with trigonometric terms could be fitted to intraday data in order to model the seasonality. As the seasonality in the data in this paper changes over time, there is appeal in using a weighted estimation approach, where more recent observations are assigned greater weights. An approach of this type, put forward in this journal 40 years ago [7], is exponentially weighted regression (EWR), which involves the inclusion of a single decay factor, λ, in a least squares minimization. A development of this, presented in [6], is the use of discount weighted regression (DWR), which enables each parameter to have its own distinct discount factor [13]. This allows different rates of discounting for different components of a time series.

Both EWR and DWR were implemented as described in [6]. Based on the estimation sample, signal coherence was used to select trigonometric terms [14]. The implementation employed the same signal coherence floor value of 0.45 that was used for half-hourly load data in [14]. For the British data, the procedure selected 70 pairs of sine and cosine terms. For DWR, three discount factors were used: λ_1 for the intercept; λ_2 for the trigonometric terms corresponding to the harmonics of the intraday cycle; and λ_3 for the remaining trigonometric terms. For the British data, the optimized values were: λ_1=0.977, λ_2=0.994 and λ_3=0.998. The relatively high discount factors for the trigonometric terms can be interpreted as slower evolution in the seasonality than in the level of the series. Implementing EWR for the British data delivered 0.996 for the single decay factor. Given the similarity of this value to λ_2 and λ_3 for the DWR method, it would seem the main difference between the implementations of DWR and EWR was in smoothing the non-seasonal component of the series.

E. DWR Spline

A cubic spline consists of smoothly joined cubic polynomials. The smoothness is ensured by imposing continuity of the spline function and its first and second derivatives at the joining points. The abscissa values for these points are referred to as 'knots'. A regression spline models the data as a spline function plus noise. It can be expressed as a linear model, and hence, after selecting the knot locations, the spline can be fitted using OLS regression [15]. In this way, a spline could be fitted to the intraweek cycle in the load data. However, this would require the incorrect assumption that the intraweek cycle is constant. Therefore, instead of OLS, this paper follows the approach in [6], which fits a time-varying spline using DWR.

As in [6] and [16], the number and locations of the knots were selected subjectively. Relatively more knots are needed for parts of the cycle where the slope is more rapidly changing. For each day, knots were positioned every hour, on the hour, except for 4pm. This implies 23 knots for each day of the week, plus an extra knot at the beginning of the week. For efficiency, using the procedures described in [16], the spline function was constrained to have the same value at selected knots. In view of Fig. 2, for the British data, the following constraints were imposed: the function on Wednesday and Thursday must be identical to Tuesday; the function at the first nine Friday knots must be the same as for Tuesday; and the function at the last 14 Monday knots must be the same as for Tuesday. This led to a spline with 93 distinct knots.

In [6], it is explained that separate smoothing of the level is enabled by first defining as a 'base knot' the knot at the start of each intraweek spline function. The value of the spline function at each of the other knots is then defined relative to this base knot. Three different discount factors are used: λ_1 for the base knot; λ_2 for the knots positioned at periods during the night; and λ_3 for the remaining knots. For the British data, the following values were obtained: λ_1=0.988, λ_2=0.996 and λ_3=0.995. As for DWR with trigonometric terms, the value of λ_1 is noticeably lower than the values of λ_2 and λ_3. This is consistent with the interpretation in [6] that using a distinct discount factor for the base knot enables smoothing of the non-seasonal component of the time series.

F. Spline-Based Exponential Smoothing

An alternative approach, based on a time-varying spline, is to use a multiple source of error state space model, with a state specified for the value of the spline at each knot [16]. This paper uses the similar approach in [6], which involves an exponential smoothing model, with the same sets of knots that that were used for the DWR spline.

G. SVD-Based Exponential Smoothing

Singular value decomposition (SVD) enables a multivariate dataset to be reduced to a dataset of lower dimension consisting of uncorrelated variables, which capture most of the variation in the original dataset. SVD has been used to refine the set of inputs to a neural network load forecasting approach [10], and within pre-processing prior to electricity price forecasting [17]. For our application, the original 'multivariate' dataset is the intraday data arranged as a ($w\times m_2$) matrix $\boldsymbol{Y}$, where w is the number of weeks in the time series. Each column of this matrix contains the observations for a particular half-hour of the week. Forecasting the next intraweek cycle amounts to forecasting the next row of the matrix. Applying SVD simplifies the task to one of forecasting the next row of a matrix with fewer columns.

The SVD of the data matrix $\boldsymbol{Y}$ delivers $\boldsymbol{Y}'\boldsymbol{Y} = \boldsymbol{VSV}'$, where $\boldsymbol{V}$ is an ($m_2\times m_2$) matrix. The columns of $\boldsymbol{V}$ are orthogonal basis functions, which can be referred to as 'intraweek feature vectors'. $\boldsymbol{S}$ is a diagonal matrix with positive entries. The 'singular values' are the positive square roots of these entries. They are placed in decreasing order, and the columns of $\boldsymbol{V}$ are correspondingly rearranged. Projecting the weekly profiles (rows of $\boldsymbol{Y}$) onto the basis functions gives 'interweek feature series', which can be viewed as columns of a ($w\times m_2$) matrix $\boldsymbol{P}$, where $\boldsymbol{P} = \boldsymbol{YV}$. Therefore, each intraweek feature vector (column of $\boldsymbol{V}$) has a corresponding interweek feature series (column of $\boldsymbol{P}$). Dimensionality is reduced by eliminating all but the first k columns of $\boldsymbol{V}$ and $\boldsymbol{P}$, leaving feature vectors and series corresponding to the largest k singular values. A forecasting method can then be used for each feature series. The resulting forecasts would be projected onto the $\boldsymbol{Y}$ space to give predictions for the original variable. In this paper, k was chosen by evaluating forecast accuracy for the final six months of the in-sample data. This led to k=31 and 38 for the British and French series, respectively. There is a link between SVD and principal component analysis (PCA) [18],[19]. PCA involves the application of SVD to a column-centered matrix. The columns of matrix $\boldsymbol{P}$ are then the principal components.

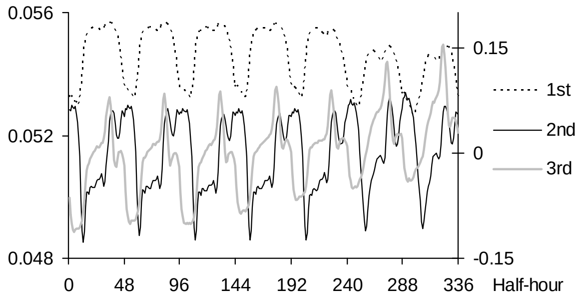

Fig. 3. Intraweek feature vectors for the British log load series. Only the 1st feature vector is plotted on the primary *y*-axis.

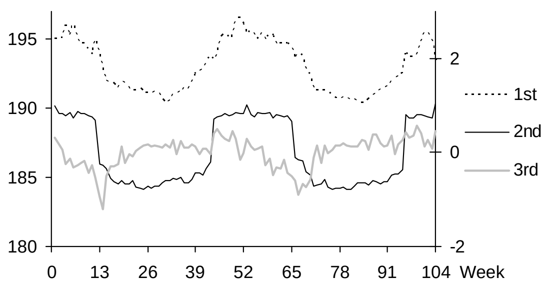

Fig. 4. Interweek feature series for the British log load series. Only the 1st feature series is plotted on the primary *y*-axis.

Figs. 3 and 4 plot the first three feature vectors and series for the British log load series. The first intraweek feature vector would seem to be the average intraweek cycle, and the corresponding interweek feature series would seem to capture the annual cycle. The second and third intraweek feature vectors show different intraweek patterns, and the second and third interweek series show how these different patterns impact to varying degrees throughout the year. The change in the intraweek pattern over the year was highlighted in Fig. 1.

From a search of the literature, it would seem that no previous studies with intraday data have applied SVD to a ($w\times m_2$) data matrix. Although SVD is applied to intraday data in [5], [6] and [19], in these studies, the data is arranged as a ($d\times m_1$) matrix, where d is the number of days in the time series. The columns of this matrix are daily series, which possess weekly seasonality. The same is true of the columns of the matrix $\boldsymbol{P}$ resulting from SVD, i.e. the feature series. In [5] and [19], simple seasonal time series methods are independently applied to each of these k feature series. Forecast updating occurs when a complete new seasonal cycle (i.e. row of the data matrix) has been observed. To enable within-day updating, an additional stage of modeling is used in [5] and [19], which increases the complexity of these methods. An appealing aspect of the SVD-based exponential smoothing method in [6] is that all k feature series are modeled together, with the result that forecasts are updated as load for each new half-hour period is observed. This model requires days of the week with similar patterns of load to be treated as identical, as in IC exponential smoothing. This brings an unappealing element of subjectivity into the method.

The length of the series in this paper allows the application of SVD to the data arranged as a ($w\times m_2$) matrix. For the resulting intraweek feature series, an exponential smoothing model is used of similar type to the one in [6]. However, relative to that model, the new model has the advantage that no subjective grouping of days is needed, as the SVD extracts from the intraweek cycle the main underlying features amongst the periods of the week. Furthermore, working with a ($w\times m_2$) data matrix leads to a simpler model formulation. The new SVD-based exponential smoothing model is presented in (6)-(8). The model updates estimates of the first k interweek feature series (columns of $\boldsymbol{P}$), and then projects them onto $\boldsymbol{Y}$ space to produce forecasts.

$$y_t = \boldsymbol{p}_{t-1}\tilde{\boldsymbol{V}}_{[t \bmod m_2]}{}' + \phi e_{t-1} + \varepsilon_t \tag{6}$$

$$e_t = y_t - \boldsymbol{p}_{t-1}\tilde{\boldsymbol{V}}_{[t \bmod m_2]}{}' \tag{7}$$

$$\boldsymbol{p}_t = \boldsymbol{p}_{t-1} + \left(\alpha \boldsymbol{1}_{m_2}\tilde{\boldsymbol{V}} + \delta\sum_{j=1}^{7}\tilde{\boldsymbol{V}}_{[t \bmod m_1]+(j-1)m_1} + \omega\tilde{\boldsymbol{V}}_{[t \bmod m_2]}\right)e_t \tag{8}$$

$\varepsilon_t \sim \mathrm{N}(0,\sigma^2)$; $\boldsymbol{p}_t$ is a ($1\times k$) state vector representing the values in period t of the first k interweek feature series; $\tilde{\boldsymbol{V}}$ is a ($m_2\times k$) matrix consisting of the first k intraweek feature vectors (columns of $\boldsymbol{V}$); $\tilde{\boldsymbol{V}}_i$ is the ith row of $\tilde{\boldsymbol{V}}$; $\boldsymbol{1}_{m_2}$ is a ($1\times m_2$) vector of 1's; and α, δ and ω are smoothing parameters. In (6), the state vector $\boldsymbol{p}_t$ is projected onto $\boldsymbol{Y}$ space using the intraweek feature vectors, which are columns of $\boldsymbol{V}$. In (8), to enable updating of the state vector $\boldsymbol{p}_t$, the error e_t, which is a value in $\boldsymbol{Y}$ space, is projected into $\boldsymbol{P}$ space using the intraweek feature vectors. The parameters α, δ and ω play a similar role to those in the HWT method of (1)-(5). The term with α updates $\boldsymbol{p}_t$ in the same way every period; the term with δ causes each element of $\boldsymbol{p}_t$ to be updated by an amount that depends on its relationship to the period of the day on which t falls; and the term with ω causes each element of $\boldsymbol{p}_t$ to be updated by an amount that depends on its relationship to the period of the week on which t falls. The optimized parameter values for the British data were: α=0.005, δ=0.602, ω=0.418 and ϕ=0.945.

IV. ARMA and ANN Benchmark Methods

A. ARMA

The 'double seasonal' ARMA model considered in [8] and [9] was implemented in this paper. The AR and MA parts of the model each involve the product of three lag polynomials. The first is written in terms of the lag operator L, the second in terms of L^{m_1}, and the third in terms of L^{m_2}. Unit root tests gave inconclusive evidence, and so differencing was not employed. The Box-Jenkins methodology was used for model selection.

B. ANN

The empirical study also included a univariate single hidden layer feedforward ANN with a single output. Using the final six months of the estimation sample as a hold-out sample, experimentation with various degrees of differencing led us to apply the operator $(1-L^{m_1})(1-L^{m_2})$ to the log of load prior to ANN modeling. After standardizing, this variable was specified as the ANN output. To avoid multi-step ahead prediction, a separate ANN was used for each lead time. For the model for lead time h, the set of potential inputs consisted of the value of the output variable at the forecast origin and at lags: 1, 2, m_1-h, $2m_1$-h, $3m_1$-h, m_2-h, $2m_2$-h and $3m_2$-h. The hidden layer and output layer activation functions were sigmoidal and linear, respectively. Weights were estimated using least squares with backpropagation. Using one step-ahead forecast accuracy for a hold-out sample consisting of the final six months of the estimation sample, the following were selected: the input variables; the number of hidden units; backpropagation learning rate and momentum parameters [20, §7.5]; and regularization parameter [20, §9.2].

V. Empirical Comparison of Univariate Methods

Forecast accuracy was evaluated for each series and each lead time, from one half-hour up to 24 hours ahead, using mean absolute percentage error (MAPE), mean absolute error, root mean squared percentage error and root mean squared error. The rankings of the methods were similar for all the measures, and the relative performances of the methods were similar for the two series, and so only the MAPE values, averaged for the two series, are presented.

Fig. 5 plots the mean MAPE for the five methods developed in [6] and the new SVD-based method developed in this paper. None of these methods has previously been applied to load data. The figure shows that total and split exponential smoothing was relatively uncompetitive. The two spline methods performed similarly up to about four hours ahead, but beyond this, the DWR spline method was more successful than spline-based exponential smoothing. Of the two DWR methods, Fig. 5 shows that the use of trigonometric terms was more accurate than splines. Fig. 5 also shows that the new SVD-based exponential smoothing method, developed in this paper, outperformed the version presented in [6]. Indeed, this new SVD-based method produced results that were, to various degrees, better than those of the five methods from [6].

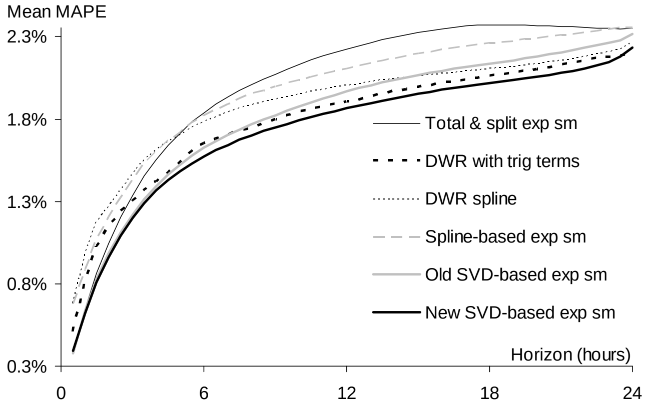

Fig. 5. Mean MAPE for the methods from [6] and the new SVD-based method.

Let us briefly consider reasons for the relative performances of the methods in Fig. 5. The total and split exponential smoothing method contrasts with the standard seasonal exponential smoothing approaches, because it essentially replaces the traditional smoothing of the level by a smoothing of the total across the periods of the seasonal cycle. Furthermore, the method assumes multiplicative seasonality, but this is not particularly appealing when, as in this paper, a log transform has been used. In comparison with the total and split method, a fundamental appeal of the DWR, spline and SVD-based approaches is that they reduce the dimensionality of the problem, and hence simplify the forecasting task [6]. With regard to the spline-based methods, they rely on suitable selections for the number and location of the knots. In this paper, this was done subjectively, but perhaps a methodical approach could be used involving cross-validation. By contrast, the new SVD-based approach requires only the selection of the number, k, of feature vectors and series, and the use of cross-validation is straightforward for this. The old SVD-based approach has the disadvantage that, in addition to the choice of k, it requires an appropriate clustering of the days of the week, as in the IC exponential smoothing method [6].

Fig. 6 presents the results for HWT exponential smoothing, the two forms of IC exponential smoothing, the ARMA and ANN benchmarks, and the EWR method with trigonometric terms. In contrast to the methods in Fig. 5, all of the methods in Fig. 6 have previously been applied to load data. Having said this, for the EWR method, in this paper, signal coherence was employed to select trigonometric terms, while in the previous use of the method the trigonometric terms were chosen subjectively [7]. The results in Fig. 6 for this EWR method are poor. Comparing these results with those for the DWR method with trigonometric terms in Fig. 5, there would seem to be a clear advantage in replacing EWR with DWR. Fig. 6 shows that the competitiveness of the ANN improved with the lead time. By contrast, the ARMA method performed well up to 12 hours ahead, but slightly less so for longer lead times. The figure shows that HWT exponential smoothing performed well, but the best results were produced by the unrestricted version of IC exponential smoothing.

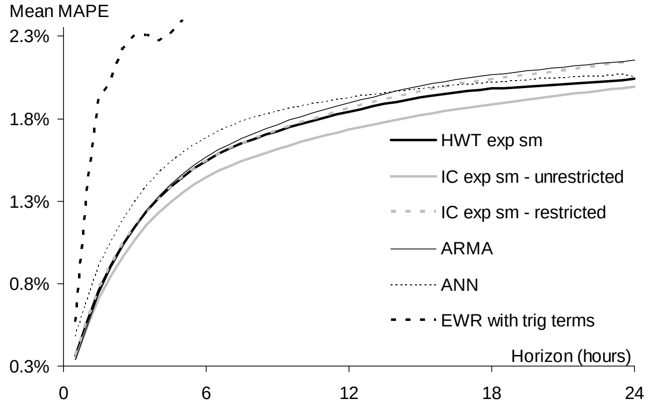

Fig. 6. Mean MAPE for the methods previously applied to load.

Fig. 7 compares the best two performing methods from Fig. 6 with the best performing method from Fig. 5, which was the new SVD-based exponential smoothing method. Also included in Fig. 7 is the SVD-based exponential smoothing method from [6]. The figure confirms that this method was less accurate than the new SVD-based method. The figure also shows that this new method performed similarly to HWT exponential smoothing, and that both were outperformed by the unrestricted form of IC exponential smoothing.

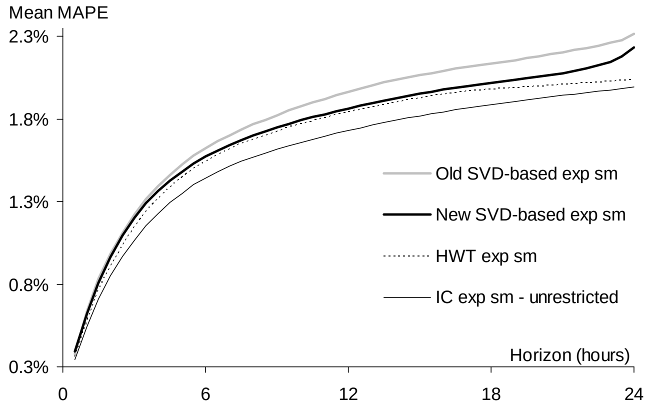

Fig. 7. Mean MAPE for the best performing methods in Figs. 5 and 6, as well as the SVD-based exponential smoothing method from [6].

Fig. 8 presents box-plots for the post-sample average percentage error (APE) for the two SVD-based exponential smoothing methods and the unrestricted form of IC exponential smoothing applied to the French data. The noticeable difference between the top two plots in Fig. 8 is in the height of the upper limits for each horizon, indicating that the largest APE was smaller for the new SVD-based method than the old SVD-based method. One cause of large errors was that the methods tended to struggle with the change in the pattern of load following the change of the clocks in October 2009. This was particularly the case for the British data. Large errors also occasionally occurred on the day following a public holiday, when the forecast origin fell on the public holiday. Although public holiday were excluded from the post-sample period, it was not felt appropriate to exclude the day following each of these days. This motivates future work looking at how best to deal with such unusual days. Although the upper limits are also high in the bottom plot in Fig. 8, the black boxes in the three plots indicate that the APE values for the IC method are less variable than for the SVD-based approaches.

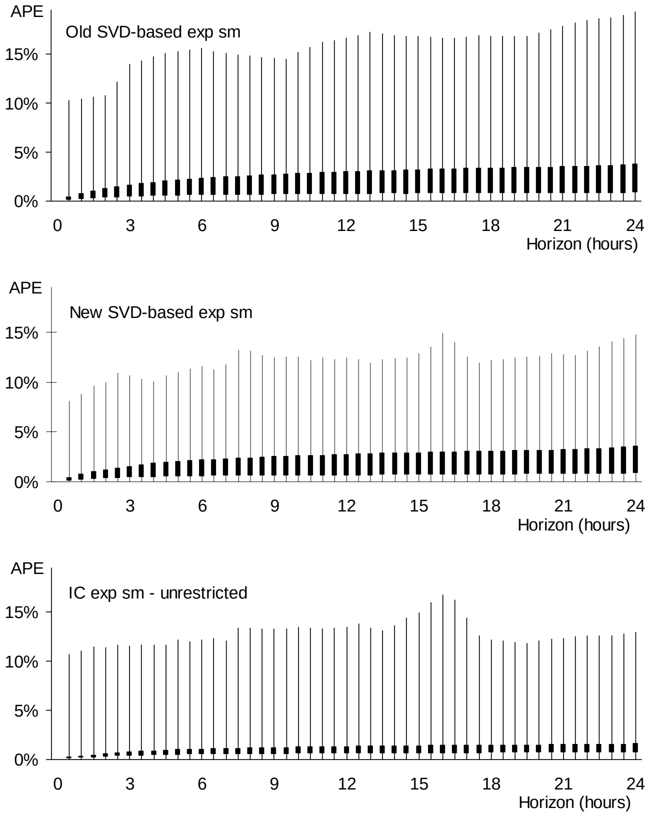

Fig. 8. APE box-plots for the SVD-based exponential smoothing methods and unrestricted IC exponential smoothing applied to the French data.

VI. Empirical Comparison of Univariate and Weather-Based Methods

This section presents an additional empirical study that compares several univariate methods with forecasts from the weather-based approach described in [8]. The empirical work in [8] employs 30 weeks of minute-by-minute British load data. This section uses the same 10-week post-sample period as that paper, with the difference that load is recorded at the half-hourly frequency. This period ran from 20 August 2006 to 28 October 2006. To be consistent with the other empirical analysis in this paper, two years of data was used for parameter estimation, rather than just 20 weeks as in [8]. The forecast origin was rolled through the post-sample period to produce forecasts from one half-hour up to 24 hours ahead.

The weather-based approach was devised by the transmission company in Great Britain, and it involves regression models estimated independently for about 10 chosen points on the daily load curve. Each model uses temperature, wind speed and cloud cover variables. Importantly, weather forecasts are used, which ensures a realistic comparison with the univariate methods. Load forecasts for the half-hours between the 10 or so chosen points are derived from a heuristic interpolation procedure involving a subjectively chosen load profile from a past day.

Fig. 9 presents the results for the weather-based approach, three univariate methods, and a combined forecast constructed as the simple average of the weather-based method and the two exponential smoothing methods included in the figure. Combining is a convenient way to synthesize the information in different individual forecasts. Indeed, it is not clear how to incorporate, in a single model, the features of the exponential smoothing and the weather-based methods. The results for IC exponential smoothing were a little poorer than those for the HWT method. In Fig. 9, the exponential smoothing methods outperform the weather-based method up to about 5 hours ahead, but beyond this the weather-based method was better. Interestingly, the combination delivered the best results at all horizons. Similar findings are reported in [8] for minute-by-minute data. This suggests that the combination efficiently synthesizes the information in the different types of forecasts.

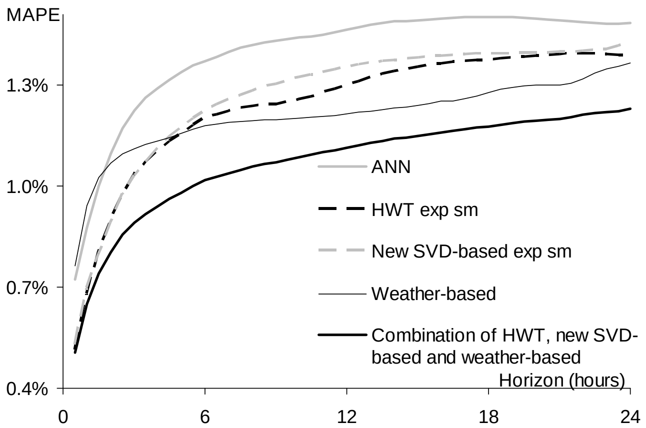

Fig. 9. MAPE for the weather-based method and several univariate methods applied to the British load data from the second empirical study in this paper.

VII. Summary and Concluding Comments

This paper has evaluated the five exponentially weighted methods developed in [6] for forecasting load up to one day ahead. The relative performances of the methods differed from the one previous empirical study with these methods in [6], where they were used for forecasting call center arrivals up to a fortnight ahead. In this paper's load forecasting application, all five methods were outperformed by the new SVD-based exponential smoothing formulation developed in this paper. In [6], SVD is essentially performed on the intraday cycle, while in the new method, SVD is applied to the intraweek cycle, which leads to a simpler and potentially more efficient model formulation. This constitutes a methodological contribution in this largely empirical paper. The better univariate methods outperformed a weather-based method up to about five hours ahead. These univariate methods were also of value for the longer lead times, because combining them with the weather-based approach led to the best results of all methods.

The methods, that had not previously been applied to load data, were not able to outperform the HWT and IC methods. Of these two, the HWT method has the appeal of simplicity. Also of practical importance is the ability to produce prediction intervals from a method. These are useful for risk management by the system operator, as well as those buying and selling electricity. For exponential smoothing models, prediction intervals can be generated using Monte Carlo simulation, but their derivation for DWR-based approaches is not obvious. In terms of future research, SVD could be used prior to weather-based load modeling, and the exponentially weighted methods could be adapted for electricity prices.

VIII. References

[1] H. Hahn, S. Meyer-Nieberg, and S. Pickl, "Electric load forecasting methods: Tools for decision making," *Eur. J. Oper. Res.*, vol. 199, pp. 902-907, 2009.

[2] D.W. Bunn, "Forecasting loads and prices in competitive power markets," *Proc. IEEE*, vol. 88, pp. 163-169, 2000.

[3] P.-F. Pai, and W.C. Hong, "Support vector machines in simulated annealing algorithms in electricity load forecasting", *Energy Convers. Manage.*, vol. 46, pp. 2669-2688, 2005.

[4] P.G. Gould, A.B. Koehler, F. Vahid-Araghi, R.D. Snyder, J.K. Ord, and R.J. Hyndman, "Forecasting time-series with multiple seasonal patterns", *Eur. J. Oper. Res.*, vol. 191, pp. 207-222, 2008.

[5] J.W. Taylor, and P.E. McSharry, "Short-term load forecasting methods: An evaluation based on European data," *IEEE Trans. Power Syst.*, vol. 22, pp. 2213-2219, 2007.

[6] J.W. Taylor, "Exponentially weighted methods for forecasting intraday time series with multiple seasonal cycles," *Int. J. Forecast.*, vol. 26, pp. 627-646, 2010.

[7] W.R. Christiaanse, "Short-term load forecasting using general exponential smoothing," *IEEE Trans. Power App. Syst.*, vol. PAS-90, pp. 900-910, 1971.

[8] J.W. Taylor, "An evaluation of methods for very short-term load forecasting using minute-by-minute British data," *Int. J. Forecast.*, vol. 24, pp. 645-658, 2008.

[9] I. Moghram, and S. Rahman, "Analysis and evaluation of five short-term load forecasting techniques," *IEEE Trans. Power Syst.*, vol. 4, pp. 1484-1491, 1989.

[10] W. Charytoniuk, and M.-S. Chen, "Very short-term load forecasting using artificial neural networks," *IEEE Trans. Power Syst.*, vol. 15, pp. 263-268, 2000.

[11] J.W. Taylor, "Triple seasonal methods for short-term load forecasting," *Eur. J. Oper. Res.*, vol. 204, pp. 139-152, 2010.

[12] J.W. Taylor, "Short-term electricity demand forecasting using double seasonal exponential smoothing," *J. Oper. Res. Soc.*, vol. 54, pp. 799-805, 2003.

[13] J.R.M. Ameen, and P.J. Harrison, "Discounted weighted estimation," *J. Forecast.*, vol. 3, pp. 285-296, 1984.

[14] J. Foster, M.J. Hinich, and P. Wild, "Randomly modulated periodic signals in Australia's national electricity market," *Energy J.*, vol. 29, pp. 105-129, 2008.

[15] D.J. Poirier, "Piecewise regression using cubic splines," *J. Amer. Stat. Assoc.*, vol. 68, pp. 515-524, 1973.

[16] A. Harvey, and S.J. Koopman, "Forecasting hourly electricity demand using time-varying splines," *J. Amer. Stat. Assoc.*, vol. 88, pp. 1228-1236, 1993.

[17] J. Chen, S.-J. Deng, and X. Huo, "Electricity price curve modeling and forecasting by manifold learning," *IEEE Trans. Power Syst.*, vol. 23, pp. 877-888, 2008.

[18] I.T. Jolliffe, *Principal Component Analysis*, 2nd edition. New York: Springer-Verlag; 2002.
[19] H. Shen, and J.Z. Huang, "Interday forecasting and intraday updating of call center arrivals," *Manufacturing and Services Operations Management*, vol. 10, pp. 391-410, 2008.
[20] C.M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, UK: Oxford University Press; 1997.

James W. Taylor is a Professor of Decision Science at the Saïd Business School, University of Oxford. His research interests include energy forecasting, exponential smoothing and density forecasting.