Short term load forecasting and the effect of temperature at the low voltage level

Stephen Haben[a,*], Georgios Giasemidis[b], Florian Ziel[c], Siddharth Arora[a,d,e]

[a] *Mathematical Institute, University of Oxford, UK*
[b] *CountingLab Ltd. & CMoHB, University of Reading, UK*
[c] *Faculty of Economics, University Duisburg-Essen, Germany*
[d] *Saïd Business School, University of Oxford, UK*
[e] *Somerville College, University of Oxford, UK*

Abstract

Short term load forecasts will play a key role in the implementation of smart electricity grids. They are required to optimise a wide range of potential network solutions on the low voltage (LV) grid, including integrating low carbon technologies (such as photovoltaics) and utilising battery storage devices. Despite the need for accurate LV level load forecasts much of the literature has focused on the individual household or building level using data from smart meters or on aggregates of such data. In this study we provide detailed analysis of several state-of-the-art methods for both point and probabilistic LV load forecasts. We evaluate the out-of-sample forecast accuracy of these methodologies on 100 real LV feeders, for horizons from one to four days ahead. In addition, we also test the effect of temperature (both actual and forecast) on the accuracy of load forecasts. We present some important results on the drivers of forecasts accuracy as well as the empirical comparison of point and probabilistic forecast measures.

Keywords: probabilistic load forecasting, low voltage networks, temperature effects, short term load forecasting

*Corresponding author
Email address: stephen.a.haben@gmail.com (Stephen Haben)

1. Introduction

Increased monitoring of the electricity distribution network through advanced metering infrastructures (such as smart meters and substation monitoring) is providing enhanced visibility and new opportunities for managing and planning the demand on the low voltage (LV) networks. This is particularly desirable for distribution network operators (DNOs) who must prepare for the increased network stressors and distributed generation as we move to a low carbon economy. Accurate load forecasts can facilitate the management of LV networks in a number of ways including: demand side response [1], storage control [2, 3], energy management systems [4, 5] and integrating distributed energy resources [6].

Low voltage networks are typically the final steps in the distribution electricity network that deliver electrical energy directly to the final customer. In the UK, and many distribution networks around the world, the final 230-240V substations contain between 1 to 6 feeders and supply electricity to between 1 and 150 customers [7]. It is worth noting that feeders are particularly heterogeneous, connected to a diverse mix of domestic customers, non-domestic customers, street furniture, overnight storage heaters and landlord lighting systems amongst many other more minor loads. The diversity of the different types of feeders is only expected to widen with the increasing uptake of low carbon technologies (LCTs), such as solar panels, electric vehicles, battery storage devices and heat pumps. The effects of particular technologies on the network can be non-trivial as we will demonstrate in this study.

In the current research, LV level forecasting has largely been focused on aggregates of smart meter data. Although smart meter aggregates can be used to describe important features of the LV network, it is important to appreciate that such aggregates are limited in their complete representativeness of LV demand, an aspect that we highlight in this study [8, 9]. The major limitation is that the aggregates are not constructed with the intention to replicate real low voltage feeders. In particular this means the resultant aggregated demand do

not consider network losses, landlord lighting, and other, typically unmonitored, network loads [7, 10]. Further, such aggregates do not explore aggregates formed of mixes of residential and commercial customers, which is a major disadvantage since many LV networks are of this form and commercial customers can have large impacts on the overall network demand. Since commercial demand is bespoke to the type of business (school, church, warehouse etc.) the random sampling procedure employed for aggregate profiles does not easily generalise to such mixed networks. A second drawback of the aggregate approach is that the relationship and natural correlations between customers cannot be taken into account through the random sampling process [8]. This has a number of important implications for the expected accuracy of LV feeder load forecasts. For example, in the UK, local residential areas can be built with similar specifications, meaning features such as overnight storage heaters (OSH) can be clustered on a single feeder. As we will show in this paper, this has a significant effect on the expected accuracy of OSH-heavy feeders' forecasts. Similar effects could be envisioned due to other significant, socially driven technologies, for example electric vehicles and photovoltaics which can be clustered due to the "keeping up with the Joneses" effect [11]. Further, customers on the same feeder are likely to respond in similar ways to local effects such as temperature, which is not possible for the randomly sampled customers in the aggregated data. Finally, even if connectivity information is taken into account, it has been shown that modelling of LV networks is less accurate via Monte Carlo methods compared to methods that take into account the substation demand data [7].

The aim of this paper to contribute to the field of low voltage level load forecasting, whereby previous studies have mostly ignored LV level load forecasting, despite their essential utility within the future smart grid applications. Moreover, we provide an exhaustive analysis on the impact of temperature estimates on load forecasting accuracy at LV level using a range of parametric and nonparametric models. The methods are applied to a large number of real LV substation feeders based in Bracknell, a town in the South of England.

The novel contributions of this study are two-fold. First, to the best of our

knowledge, it is the first large-scale study on real LV feeder data, which exhibit richer structure than aggregates of smart-meters. In contrast to aggregated data, LV networks consist of different proportions of a variety of customers (domestic, small-to-medium enterprises (SMEs) and commercial, schools, hospitals, etc.) as well as other, unmonitored street furniture such as street lighting and traffic lights [10]. Hence, there is a limit to the representativeness of the aggregated results. Second, temperature effects on LV load forecasting are explored, in particular, whether the relationship between load and temperature is causal or simply a correlation effect for real LV feeders. To this end, both actual and forecast temperature data from the region of interest are being used, which allows us to compare the difference between using historical and forecast temperature in our load forecasts[1].

Besides verifying existing results in the literature, through our analysis, we contribute with two main innovative results. First, we show that the power law relationship between size of the feeder and forecasts accuracy, [12], does not hold for particular types of feeders, e.g. those with overnight storage heaters, and/or large landlord lighting supplies. Simulated LV networks via aggregates of smart meters as presented in the literature are unable to unveil such behaviour, because they do not consider real LV network connectivity. This learning output could be used by network operators to plan, manage and operate the future networks consisting of a variety of smart technologies. Second, we find that at the LV level the temperature does not have a causal effect on the load forecast accuracy, and it is actually the strong seasonal correlation that might have a stronger effect on forecast accuracy.

The rest of the paper is organised as follows. In the next Section 2, we present a literature review on electricty demand forecasting (at all voltage levels). In Section 3, we analyse the data that we use in this study. In Section 4, we

[1] In much of the load forecasting literature, temperature forecast data is not always available and historical weather is used instead. This limits the feasibility and conclusions since such forecasts are not possible in practice.

describe the methods as well as the scoring functions we use to measure the accuracies of our forecasts. In Section 5, we describe the main results and in the final section there is a discussion including potential future work.

2. Literature Review

Load forecasting has traditionally been implemented for the high voltage (HV) or system level which typically consists of the aggregated demand of hundreds of thousands or millions of consumers. By the law of large numbers such demand is much less volatile than LV demand and hence, relatively easier to predict. Load forecasting at this level is a very mature research area and hence, there is vast literature describing and testing a variety of techniques including, artificial neural networks (ANNs), support vector machines, ARIMA, exponential smoothing, fuzzy systems, and linear regression. For a literature review of the recent methods see [13, 14] as well as the review paper for the Global Energy Forecasting Competition (GEFCom) 2012 [15]. Many of these papers have shown strong correlations between weather effects and demand[16, 17], for example, this is exhibited in [18] for a large number of European countries with the relationship dependent on the climate. In [19] the authors show the link between temperature and district heating for a region in Denmark. Due to the strong relationships between weather and load forecasting at the system/HV level, in [20] the authors included historical weather data in their short term load forecasts, and the authors in [10] considered historical values of wind chill, humidity and wind speed. In [21] the authors used weather forecasts to produce load forecasts for a large urban area in Australia. Due to the high regularity at such large aggregations the mean absolute percentage errors (MAPEs) are typically small at around $1.5 - 6\%$.

In contrast, published research on forecasting at the lower voltage levels has focused on short term load forecasts at the household level, applied to smart meter data. However, large quantities of smart meter data are not currently available and so much of the research has been restricted to those which are in

the public domain such as the Irish smart meter data set [22]. Due to the high volatility at the household level, it is particularly difficult to produce accurate forecasts. In fact, the authors in [23] showed that due to the "double peak" error for spiky data sets, it is difficult to objectively measure the accuracy of household level point forecasts using traditional pointwise errors. As with the HV level, similar methods have been applied to the household level, including ANNs [24], ARIMAs [25], wavelets [5], Kalman filters [26], and Holt-Winters exponential smoothing [27]. The errors in these methods are much larger than the HV level, with MAPEs ranging from 7% up to 85% in some cases [24, 25]. A link between weather and household demand has been observed and historical weather data has been used within the methods [27, 1]. Some of the strongest correlations observed have been in temperature and illumination [28].

The literature on LV level forecasting is sparse compared to both HV and household levels. LV distribution feeders are relatively volatile compared to HV level demand since they consist of low aggregations of customers (typically less than 150 in the UK) [7]. The main forecasting research has been presented in [6, 29] where the authors apply both ARIMAX and ANN methods to a single LV transformer (consisting of 128 customers) to forecast total energy and peak demand. They achieve MAPEs of between $6 - 12\%$. They also included historical weather data in the methods. At slightly higher voltage substations, the authors in [30] and [31] apply ANN and ARIMA methods to MV/LV level data to achieve MAPEs of 11% and $13 - 16\%$ respectively. The majority of load forecasts at the LV substation level are in fact applied to aggregates of smart meter data [32, 12, 33]. In [32, 12] the authors consider a variety of methods and show a strong scaling law relationship between the MAPEs and the aggregation size. In [12] the authors consider the aggregation level of data from 1 to 100k smart meters. The relationship between relative accuracy and aggregation is verified in [33] where the authors consider aggregations of 1, 10, 100, 1000, 10000 smart meters using ANNs applied to data from 40,000 customers in Basel, Switzerland. MAPEs vary considerably from as low as 2% [34] up to 30% [35]. The relationship between temperature and load has had mixed results at the LV

level based on aggregated data. In [35] the authors successfully apply weather data with ANN and ARMA methods to 4 different levels of aggregations (1, 10, 100, 1000 smart meters) from both the Irish and a Danish smart meter set. In [36], the authors consider aggregates of 5, 10 and 100 smart meters from the Irish data set. They do not utilise weather data as they argue that the weather data is not scalable. In [37] the authors consider two main data sets consisting of 40 time series and apply ARIMA, Holt-Winters, ANN, and generalised additive models. Although they show there is a correlation between weather and load in the data sets, they suggest that there is not much effect of temperature on load.

Due to the volatility of LV level demand, probabilistic load forecasts are a natural choice to provide a detailed description of their uncertainty. Most recently there has been an increased interest in probabilistic load forecasting, accelerated by the recent Global Energy Forecasting Competition 2014 [38]. See [39] for an up to date review of the current state of the art and major challenges in probabilistic load forecasting. There have been a few publications where the authors have considered load forecasting of individual smart meter data or on aggregations of smart meter data. In [40] the authors applied kernel density estimation methods to the Irish smart meter data and compared the errors between the forecasts of domestic and non-domestic customers over a horizon of up to a week ahead. In [41], probabilistic forecasts of smart meter data from 226 Portuguese households ware considered using quantile regression with a generalised additive model (GAM). In [42] the authors also use a GAM considering both individual and aggregations of the Irish smart meter data up to 1000 smart meters. They found at large aggregations that a normal distribution was a sufficient model for the demand, consistent with what we would expect from the law of large numbers. This is in contrast to LV level demand which is not typically normally distributed. The same authors also considered various aggregation levels of UK smart meter data in [8] using copulas to develop the joint distributions at different levels of aggregation and ensure coherent probabilistic forecasts. Finally, in [43] day ahead quantile forecasts are created using Laplace distributions and non-parametric methods.

In the research presented here we consider both point and probabilistic forecasts and compare both point scoring functions (such as MAPE) to probabilistic scoring functions (continuous ranked probability score). We will show that it may be sufficient to evaluate accurate point forecasts to produce corresponding accurate probabilistic forecasts. This supports the research considered in [44] which compared the use of MAPE and the pinball score for selecting the parameters for a probabilistic model.

3. Data Analysis

In this section we review and analyse the data that will be used to create and evaluate our methods.

3.1. Load Data

The load data for 100 feeders begins on 20^{th} March 2014 up to the 22^{nd} November 2015, a total of 612 days. The feeders consist of a range of magnitudes with the average daily demand of approximately 602kWh and a maximum and minimum daily demand of around 1871kWh and 107kWh respectively. Out of the 100 feeders, 83 are connected purely to residential consumers with an average of 45 households, a maximum of 109 customers and a minimum of 8. A further 7 have no connectivity information. For the trial presented in this paper we define a test set, consisting of the dates 1^{st} October 2015 to 22^{nd} November 2015 inclusive. The remainder of the data is used for training.

The data contains strong intraday and intraweek seasonality. Figure 1 shows the relationship between the autocorrelation at lag 168 (i.e. a week) and the mean daily feeder demand. The plot highlights that all feeders have some degree of weekly regularity with the larger feeders tending to have much stronger autocorrelation than smaller feeders, which indicates that the larger feeders will tend to have lower forecast errors.

There is also a strong seasonal effect in the data, however note, that due to the lack of air conditioning in the UK there is no increase in demand for

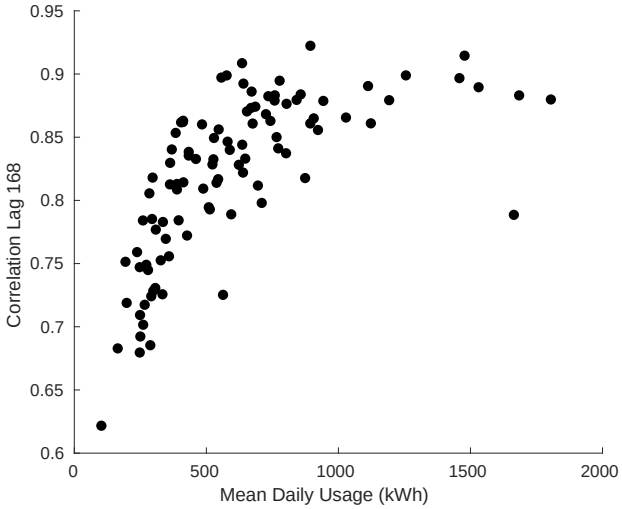


Figure 1: Autocorrelation at lag 168 (intraweek seasonal cycle length) for all 100 feeders against the mean daily demand.

warmer periods in contrast to other data sets which are commonly shown in the literature [38].

3.2. Weather Forecasts

Weather variables, especially temperature, and those related to temperature such as wind chill, are often included within load forecasting models [45, 18]. Typically those are for high voltage load forecasts and hence represent the demand of a large number of customers. However, it is not obvious that weather variables play an important role in load forecasts at the LV level. In this study we consider temperature data (in degrees), collected from the Farnborough weather station, the closest weather station to Bracknell at just under 10 miles (approx 16 Km). In our forecasts we will utilise both the historical hourly and forecast temperature data[2].

Although weather has a strong correlation with demand at the high voltage,

[2]The weather forecasts have been provided by Meteogroup.

for practical load forecasts, only temperature forecasts can be used. We first consider the accuracy of such forecasts before considering the correlation of temperature on the LV load forecasts. The forecasts begin at 7am on each day and forecast each hour for the next four days. In other words, the one hour ahead forecasts are all for the period 8am, the two hour ahead forecasts are all at 9am, etc. Thus, the temperature forecast accuracy is not simply determined by the forecast horizon (where usually greater forecast horizons correspond to more inaccurate forecasts), but also the volatility of the time period being estimated. Indeed the accuracy (as measured by the mean absolute percentage error) does reduce as a function of horizon, as shown in Table 1, for the full data set. Forecasting temperature four days ahead (i.e. between 73 and 96 hours ahead) has dropped in accuracy by around 100% compared to up to one day ahead (between 1 and 24 hours ahead). For the test set the temperature forecast accuracy is improved compared to the training set (see Table 5 in Section 5.4).

Forecast Horizon	1 Day	2 Day	3 Day	4 Day
MAPE	11.85	15.60	20.21	23.80

Table 1: MAPE for the temperature forecasts for different horizon periods (over the period 31^{st} March 2014 to 28^{th} Nov 2015).

In our analysis we will focus on the relationship between the load forecasts using forecast temperature inputs (ex-ante forecasts) rather than the historical temperature inputs (ex-post forecasts) [46]. Although the forecast and actual temperature values are very strongly correlated (with a pairwise linear correlation coefficient greater than 0.95) we will include results for both ex-ante and ex-post versions. We do this for two reasons. Firstly, much of the literature presents so-called "ex-post" forecasts (i.e. those which use the historical temperature inputs) and hence comparisons are possible with other research [46]. Secondly, it provides further evidence concerning the causal link between temperature and load at the LV level.

As with high voltage load, there are correlations, usually negative, between temperature and the LV loads from our feeders. It was found that individual hours of the day typically had, on average, stronger correlations with the temperature than the correlations between the complete temperature and load time series. In the Appendix A, we present an example of a feeder from the trial with a relatively strong correlation between load and temperature, showing the linear relationship which varies depending on the hour of the day. This suggests that it may be worthwhile considering splitting the time series into 24 separate, hourly time series for some of our forecasts.

This section has presented some basic analysis of the load data and corresponding temperature data. We have highlighted some important features which should be utilised in the load forecasts. Daily, weekly and annual periodicities have strong relationships to load demand and there are different behaviours in demands for different periods of the day. Further, we have seen there is a wide variety of feeders (in terms of numbers of connected customers and magnitudes of demands) with different correlations with the temperature. We will use this information to construct our forecasts as well as appropriate benchmarks.

4. Methods

In this section we present a wide variety of methods which will be used to generate the forecasts. A number of benchmarks are also included for comparison as well as to highlight the importance of various inputs/predictors.

The chosen methods are motivated by the data analysis presented in Section 3 and are appropriate for volatile data for a number of reasons. First, the kernel density (Section 4.1) and quantile regression (Section4.2) methods have been proven successful in the Global Energy Forecasting Competition (GEFCom) 2014 [38]. Second, these methods have been successfully applied to more volatile datasets, e.g. the kernel density and the Holt-Winters-Taylor method (Section 4.4) have accurately forecast smart-meter data [40], which are more volatile than LV demand, whereas the autoregressive methods (Section 4.3)

have been applied to volatile financial time-series [47]. Third, the simple seasonal regression (Section 4.2) and the autoregressive models have been used to accurately forecast LV demand data for LV feeder storage applications [48]. For these reasons, we believe, that these models are appropriate for volatile data and span a wide range of methods. Further techniques, such as machine learning (e.g. neural networks, support vector machines, etc.) lie beyond the scope of this empirical study and will be the focus of future studies.

For this section, without loss of generality, we define L_t, $t = 1, ..., N = D \cdot H$ to be the hourly time series for the load for a particular feeder, where $H = 24$ and $D = 612$ is the number of days in the training and test data set combined. The initial time step $t = 1$ defines the start of the data set 20^{th} March 2014 for the hourly period 12AM to 1AM. Also, without loss of generality, we also define t_h to denote the end of the historical data (with $t_h + 1$ therefore the start of the testing period) which determines the maximum available training data for the methods. Specifics of each methods will be described in their corresponding sections.

4.1. Kernel Density Estimation

The first method we consider are those based on kernel density estimation (KDE) techniques. These have been successful at generating probabilistic forecasts for individual smart meter data as well as in the GEFCom 2014 competition on higher voltage demand [40, 49]. The method aims to generate a conditional distribution at time t, $f(L_t|X)$, conditional on historical demand and other factors, X, such as the period of the week or weather variables. The main challenge of such methods is the computational expense of evaluating the correct parameters, in particular the bandwidths. In this section we presume the parameters are trained on the historical data $[t_1, t_2]$, with $1 \leq t_1 < t_2 \leq t_h$, $t_1 \mod H = 1$ and $t_2 \mod H = 0$, in other words the period of time encapsulates full days from the historical data.

We consider four variations of the KDE methods. The first type, denoted as **KDE-W**, is the unconditional KDE trained on all observations from the same

hour of the week (1 to 168). The second type, referred to as **KDE-Wλ**, includes a weighting parameter that favours observations around the same period of the year as the forecasting time, see Appendix B and [49] for further details.

Additionally, we consider kernel density estimate forecasts conditioned on independent variables y, z (CKD), such as the week-period, the temperature, or both. We produce three CKD forecasts, one conditioned on the week period (**CKD-W**), a second conditioned on the week period and the actual temperature readings (**CKD-WTa**), and a third forecast conditioned on the both the week period and the forecast temperature (**CKD-WTf**). For CKD-W, $y_i = i \mod 7H$ is the week period of time interval i. CKD-W weighs observations towards similar times of the week as the forecast time-period. For further details on mathematical formulas, training and parameter optimisation, see Appendix B.

An advantage of the kernel-density methods is that the entire distribution is found simultaneously whereas quantile regression methods only find individual quantiles. The disadvantage is the computational costs, especially as more conditional variables are introduced. Various methods, such as clustering the time series, must be employed to reduce the costs [40]. We consider the methods as both probabilistic forecasts and also as point forecasts by using the median of the distributions as our point estimate.

4.2. Simple Seasonal Linear Regression

The method is based on an update of the simple seasonal model presented in [49]. For this method rather than construct a full probability density function for the load distribution we instead develop models for a number of predefined quaniltes $\tau \in (0, 1)$. Here, we assume a linear regression model for each quantile, treating each period of the week as a separate time series of the form

$$\begin{aligned} \hat{L}_t^\tau &= \sum_{k=1}^{H} \mathcal{D}_k(t) \left(a_k^\tau + b_k^\tau \eta(t) + \sum_{p=1}^{P} (c_{k,p}^\tau) \sin\left(\frac{2\pi p \eta(t)}{365}\right) + (d_{k,p}^\tau) \cos\left(\frac{2\pi p \eta(t)}{365}\right) \right) \\ &\quad + \sum_{l=1}^{7H} f_l^\tau \mathcal{W}_l(t). \end{aligned} \tag{1}$$

Here, $\eta(k) = \lfloor \frac{t}{H} \rfloor + 1$ is the day of the trial (with day 1 as 20^{th} March 2014). There are two dummy variables identifying the period of the day, $\mathcal{D}_k(t)$, and the period of the week, $\mathcal{W}_l(t)$. Further details of the model can be found in the Appendix C.

For the point forecast as with the KDE methods we considered the median quantile and also a least squares estimate. In all cases the median outperforms the least squares estimate and hence our point forecast estimates will be derived from the median quantile. The methods are trained on the entire available historical information using the latest data for the rolling forecasts at the start of each new day. Further variants of the model were also considered (different numbers of seasonal terms $P = 2, 3$, with and without trend (i.e. $b_k \equiv 0, \forall k$), and also using only weekend dummy variables instead of dummy variables for all days of the week), but we found that the best methods used a linear trend, three seasonal terms ($P = 3$), and used the day-of-the-week dummy variable as in equation (1). We also found that the model without trend also performed reasonably well, so in our analysis we will consider both methods. We will denote the seasonal method with trend as **ST** and without trend as **SnT**.

Including temperature effects in the model is straightforward and only requires adding a polynomial to the full equation (in our case up to only cubic order). Depending on the horizon (one, two, three, or four days ahead) four different models are calibrated.

4.3. Autoregressive Methods

The models in this section are all based on a regression of a residual time series $r_t = L_t - \mu_t$, for some *mean* profile μ_t which we define later. This residual regression can be written,

$$r_k = \sum_{k=1}^{p_{\max}} \phi_k(r_{t-k}) + \epsilon_t \tag{2}$$

for some Gaussian error ϵ_t. The main tunable parameter in the model is the optimal autoregressive order $p = p_{\max}$ which is chosen by minimising the Akaike information criterion (AIC) over $p \in \{0, \ldots, p_{\max}\}$, then the coefficients

$\phi_1, \ldots, \phi_{\max}$ can then be easily determined by the Burg method. The methods are trained over a year's worth of historical data prior to the start of the test period on the 1^{st} October 2015.

For the mean profile μ_t we consider two models, the details of which can be found in the Appendix D. The first uses an average weekly profile which we denote as **ARWD**. The second updates the simple weekly average profile to include an annual seasonality. We denote this model by **ARWDY** .

As with the ST methods it is trivial to include weather effects by adding linear terms to the mean equations. We note that separate models are used depending on whether the forecasts are one, two, three or four days ahead.

The methods can be updated to generate probabilistic forecast methods by also modelling a variance term using the same features as the mean equation. See Appendix D for further details.

4.4. HWT Exponential Smoothing Method

We implement the Holt-Winters-Taylor (denoted **HWT**) exponential smoothing method [50] to model the intraday and intraweek seasonality in feeder load. This method is represented as:

$$\begin{aligned} L_t &= l_{t-1} + d_{t-s_1} + w_{t-s_2} + \phi e_{t-1} + \epsilon_t \\ e_t &= L_t - (l_{t-1} + d_{t-s_1} + w_{t-s_2}) \\ l_t &= l_{t-1} + \lambda e_t \\ d_t &= d_{t-s_1} + \delta e_t \\ w_t &= w_{t-s_2} + \omega e_t, \end{aligned} \tag{3}$$

where L_t denotes the load observed at time t, l_t denotes the level, $\epsilon_t \sim \text{IID}(0,\sigma^2)$, $s_1 = 24$, $s_2 = 168$, and d_t and w_t correspond to the intraday and intraweek seasonal indexes, respectively. This model requires the estimation of three smoothing parameters λ, δ and ω for the level and two seasonal indexes, along with a parameter ϕ to adjust for first order auto-correlation in the error (denoted by e_t). The model parameters were estimated separately for each feeder.

4.5. Benchmark Methods

As a comparison to the methods presented above, we implemented the following four simple benchmarks to model load for each feeder:

Benchmark 1, Last Day (LD): $\hat{L}_{t+k} = L_{t+k-s_1}$

Benchmark 2, Last Week (LW): $\hat{L}_{t+k} = L_{t+k-s_2}$

Benchmark 3, Last Year (LY): $\hat{L}_{t+k} = L_{t+k-s_3}$

Benchmark 4, Simple Average (SMA): $\hat{L}_{t+k} = \frac{1}{p}\sum_{i=1}^{p} L_{t+k-i\times s_2}$

where $\hat{L}_{t+k}$ is the k-step ahead prediction, t is the forecast origin, while $s_1 = 24$, $s_2 = 168$, $s_3 = 52 \times s_2$ denote the intraday, intraweek and intrayear cycle lengths, respectively. For benchmark 4, we considered a variety of different p values and the best in-sample results were obtained for $p = 4, 5$ weeks, which have very similar scores. We will only present results for $p = 5$. For further details regarding these benchmarks methods, please see [51].

4.6. Empirical Estimate

The aforementioned benchmark methods only provide point estimates and hence cannot provide quality comparisons to probabilistic forecasts. For each period of the week we define an empirical distribution function using all the load data from the same time period over the final year of the historical data. We then use this empirical distribution to define the desired quantiles. The median quantile is used as a point estimate. The estimate is fixed over the entire test period. We refer to this method as the **Empirical** forecast. Note that we could have included more historical data in the construction of the empirical distribution but by restricting it to the last year we do not introduce seasonal biases.

4.7. Error Measures

To evaluate our methods we consider a variety of forecast measures which are common to the forecasting and in particular, load forecasting literature. For

the point forecasts we use two common measures, the mean absolute percentage error (MAPE) and the mean absolute error (MAE).

For the probabilistic forecasts we use the continuous ranked probability score (CRPS), which quantifies both the calibration and sharpness of the forecasts [52]. Suppose we have actual loads $\mathbf{a} = (a_1, a_2, \ldots, a_n)^T$ for time periods $\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_n$ and a probabilistic forecast, given by a cumulative distribution $F_k(\cdot)$ at time $\hat{t}_k$, then the CRPS is defined by

$$CRPS(\mathbf{a}, F) = \frac{1}{n} \sum_{k=1}^{n} \int_{-\infty}^{\infty} (F_k(z_k) - \mathbf{1}(z_k - a_k))^2 dz_k, \tag{4}$$

where $\mathbf{1}$ is the Heaviside step function. The MAE and CRPS are scale dependent and therefore we cannot compare feeders of different fundamental sizes. We therefore normalise the score by dividing the MAE and CRPS by the average hourly load of the feeder over the last year of training data. We also multiply the errors by 100 so that they may be referred to as percentages. We refer to these as the relative MAE (RMAE) and relative CRPS (RCRPS) respectively. The CRPS reduces to the MAE in the case of a point forecast and hence we expect them to be strongly related [52]. The MAPE scales each error according to the size of the actual and hence needs no adjustment. The potential disadvantage of this method is that a few small loads ($a_k << 1$) which are poorly estimated could skew the average errors.

5. Results

In this section we compare the methods we have developed in Section 4. The test period is the 53 days consisting of 1^{st} October 2015 to 22^{nd} November 2015 inclusive. Since the temperature forecast data is available from hourly horizons up to 4 days (96 horizons) we will consider up to four day ahead forecasts even when not considering the temperature variables. As described in the Methods section we will construct both point and probabilistic forecasts which will be evaluated using MAPE/RMAE and RCRPS respectively. The test period does not have any holiday dates but does contain a daylight savings change-over on

the 25^{th} October 2015. However, this date will not be treated specially in this trial. Special days (public holidays) effects will be considered in future work.

5.1. Average Errors

To begin, we consider forecasting techniques without temperature inputs, this will be considered in much more detail in Section 5.4. The average error scores for the methods are shown in Table 2, these consist of the errors for all four day ahead forecasts over the test period. We do not show the LD or LY benchmarks as they are much worse than the Empirical, LW and SMA benchmarks. From comparing the benchmarks we find that the simple average SMA is the best methods for point forecasts and is, in fact, quite competitive with some of the presented methods. In fact it only has an average MAPE of 7% worse than the best methods score, ARWDY. The LY method performs the worst and there is a slight improvement by using the yesterday-as-today estimate LD. The same period of the week forecast, LY, is the best of the random walk methods and shows the strong weekly periodicity of the data.

The most accurate methods for point forecasts are ARWD, ARWDY and HWT which all include daily, weekly and annual periodicities as well as strong autoregressive components. The ARWD and AWRDY methods generate the best forecasts with the ARWD being slightly better in terms of the RMAE. The HWT is the next best method and has a MAPE of only 1.3% larger than the ARWD/ARWDY methods. The ST and SnT methods is similar to the AWRDY method but without the autoregressive components and although improves on the benchmarks, only outperforms the SMA by 2%. On average the KDE methods perform the worst only outperforming the random walk methods, LD, LW and LY. Conditioning the KDE forecasts on the week period CKD-W improves the forecasts, especially with respect to the RMAE.

For the probabilistic forecasts the empirical benchmark is also shown to be effective and gives a RCRPS only 20% worse than the ARWD method. The ARWD is once again the best forecast, closely followed by ARWDY. This time the ST forecast performs slightly better than the HWT forecast. Of the 100

feeders, the ARWD and ARWDY forecasts were the best performing, with the smallest average errors, for 16 and 26 of the feeders respectively. Hence, there does not appear to be a one-size-fits-all forecast which performs best for all feeders and identifying indicators of which forecast to choose will be an important step for practitioners.

	Error Scores %		
Method	MAPE (std)	RMAE (std)	RCRPS (std)
LW	18.67 (5.85)	18.93 (5.34)	-
SMA	15.73 (5.05)	16.77 (5.03)	-
Empirical	16.19 (5.19)	16.96 (4.95)	12.62 (4.06)
HWT	14.84 (4.60)	15.01 (4.14)	11.06 (3.03)
ARWD	**14.65 (4.71)**	**14.67 (4.12)**	**10.32 (2.85)**
ARWDY	**14.64 (4.55)**	**14.80 (4.15)**	**10.44 (2.88)**
ST	15.42 (5.20)	15.42 (4.75)	10.97 (3.31)
SnT	15.66 (5.13)	15.57 (4.70)	11.10 (3.31)
KDE-W	17.05 (5.56)	19.36 (6.80)	13.79 (4.52)
KDE-Wλ	17.08 (5.58)	19.41 (6.83)	13.80 (4.52)
CKD-W	16.54 (6.61)	17.22 (6.25)	13.23 (4.75)

Table 2: MAPEs, RMAEs and RCRPSs for all forecast methods over all 4 day-ahead forecasts for the entire 53 day test period. The lowest errors for each error score are highlighted in bold. Standard deviations are indicated in the brackets.

If we compare the average errors for each feeder for a particular method we

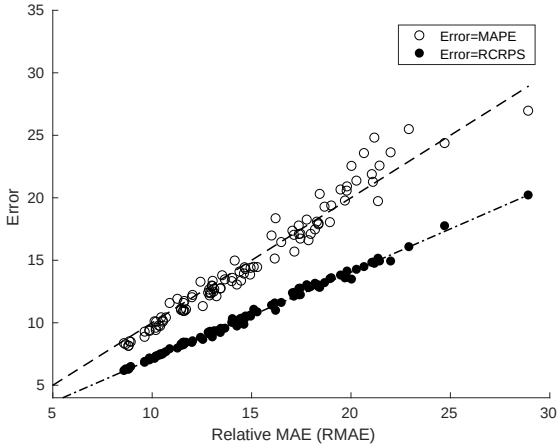


Figure 2: Scatter plot showing the average RCRPS (filled) and average MAPE (unfilled) versus average RMAE for each feeder. Also shown are lines of best fit. These results are for the ARWDY method.

see that the errors are strongly correlated. Figure 2 shows this comparison for the ARWDY method. The plots are very similar for all methods. As expected, the RCRPS and RMAE are strongly related and this corresponds to a very strong linear correlation in the average errors (0.995). The MAPE and RMAE are also strongly correlated (0.981) but with more scatter, especially for larger errors. The strong correlation between the error measures means it is inefficient to present the remaining results in terms of all scores, MAPE, RMAE, and RCRPS. For this reason, and because of the ubiquitous use in the load forecasting community we will frame the rest of our discussion and analysis with respect to MAPE.

5.2. Forecast Accuracy and Horizon

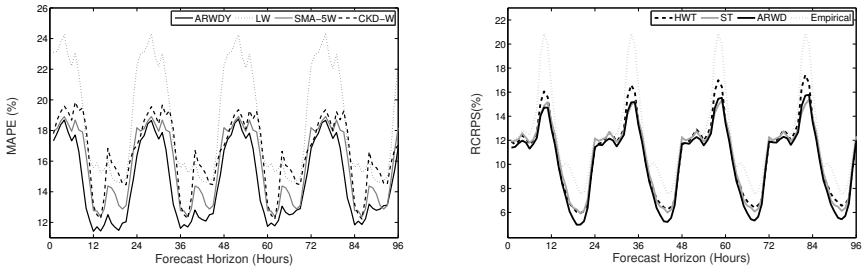
Here we investigate the drop in forecast accuracy as a function of horizon. Initially we consider the forecast accuracy in terms of full days ahead. In other words, we consider the accuracy of forecasts up to 1 day ahead, forecasts between 1 and 2 days, etc. The MAPEs as a function of whole days are shown in Table 3

for selected methods. Specifically, we generate forecasts for horizons at hourly intervals, ranging from one hour up to four days ahead. We use as forecast origin each 7am in the post-sample data, so as to be consistent with the temperature forecast data obtained from the weather station. We note that the benchmarks do not change over the 4 day horizon since they are performed a week in advance. As expected, the most accurate forecasts horizon is one day ahead and the least is 4 days ahead. However, the drop in average accuracy is quite small, with no more than a 4% drop in forecast score.

Method	MAPE			
	Day 1	Day 2	Day3	Day4
HWT	14.56 (4.46)	14.83 (4.59)	14.95 (4.67)	15.04 (4.72)
ARWDY	14.34 (4.45)	14.59 (4.52)	14.75 (4.62)	14.87 (4.65)
ST	15.36 (5.15)	15.41 (5.19)	15.44 (5.22)	15.49 (5.24)
CKD-W	16.64 (6.67)	16.56 (6.60)	16.48 (6.57)	16.50 (6.50)

Table 3: MAPE Scores for each method over each day ahead horizon. Standard deviations are indicated in brackets.

The MAPEs as a function of horizon at the hourly resolution for selected methods are shown in Figure 3a. First recall that the first horizon corresponds to the period $8-9$AM. There are a number of interesting observations. It is clear that all forecast methods produce a similar shape and the more accurate forecasts have smaller errors at all horizons. Secondly, as confirmed with Table 3 there is a general trend with a small overall increase in the error as a function of the horizon. However, the strongest driver of the forecast accuracy is clearly the period of the day. The most inaccurately forecast time periods corresponds to the hours from around 6AM to 6PM. Similarly the period most easily estimated is the evening and night period. Surprisingly, we would expect the evening



(a) MAPE (b) Normalised CRPS

Figure 3: Plot of average scores for selected methods for horizons from 1 hour to 96.

period (from 6PM until 11PM) to be quite difficult to forecast. In fact, we discover that the horizon-error shape may be an artifact of the error measure used. In Figure 3b we show the same plot but this time for the relative CRPS score for selected methods. This shows the expected larger error in the evening period. The difference between the two scoring functions is that the MAPE (4) normalises each hourly forecast error with respect to the demand at the same hour. Since for residential feeders (which dominate the composition of the feeders in this trial) have largest demand in this time period the relative error MAPEs are smaller compared to the RCRPS or RMAE. This could have important implications for the use of MAPE in predicting daily errors, especially for the many applications where peak demand is of the most importance, such as in peak demand reduction via storage devices [3]. As evident from Figures 3a and 3b, the forecast errors are high for periods of the day that witness a relatively large change in consumption. We note that since all forecasts start at 7AM each day we have no information on the accuracy of the forecast models as a function of horizon (and regardless of starting point). Encouragingly, the horizon plot shape shown in Figure 3b is consistent with that as shown by the authors in [40] for 800 individual residential customers.

5.3. Accuracy and Feeder Size

As described in the data analysis section the connectivity of the feeders considered varies and consists of different mixtures (domestic and non-domestic) and numbers of customers. Recent literature has shown there is a link between the size of the aggregation which makes up a demand time series and the accuracy of the forecast on a number of different load time series data sets [32, 12]. In contrast to the previous work, these are real networks and hence may have different behaviour compared to aggregations of smart meter data. These networks include street furniture, such as street-lighting and traffic lights, LCTs, overnight storage heaters, and correlations between neighours are also naturally incorporated.

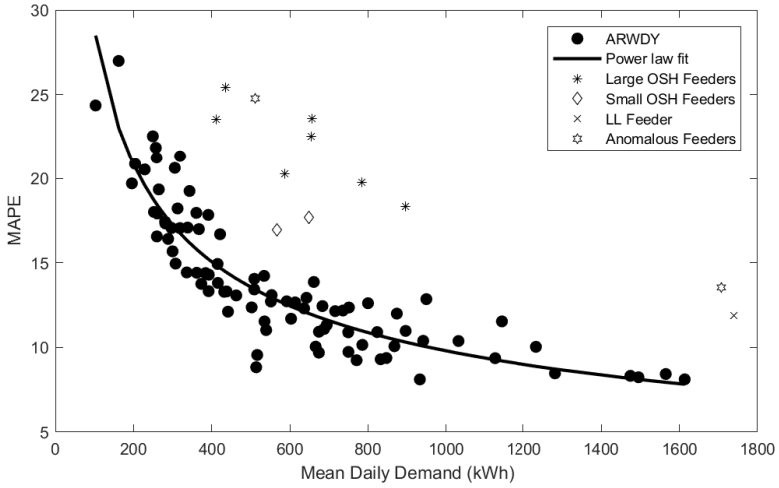


Figure 4: Scatter plot of the relationship between MAPE and mean daily demand for two different forecasting methods. Feeders with unexpectedly large errors have been labeled separately with their known connectivity characteristics (if known). Also shown is a power law fit to the non-anomalous feeders.

Figure 4 shows the MAPEs for each individual feeder as a function of the average daily demand as well. The majority of the feeders appear to fit a power law relationship which fits with the results found in [32, 12]. However, it is clear

that twelve of the feeders do not fit the relationship as tightly as the remaining 88 feeders.

After investigation it was found that these particular LV feeders consisted of different connected customers. Firstly, it was found that seven of these feeders consisted of unusually large overnight demands most likely due to overnight storage heaters (OSH). In the UK such customers are typically on a special tariff (Economy 7) which provides cheap electricity overnight. Indeed, examination of the connectivity found that 75-85% of customers on each of these seven feeders were on the Economy 7 tariff. These feeders are labelled "Large OSH Feeders" in the figure. Further, two other feeders with smaller overnight demands (but still with noticeable peaks) were discovered and found to have 62% and 75% of their customers on the Economy 7 tariff. These are labelled "Small OSH Feeders" in Figure 4. No other feeders in the 100 were found to have large overnight demands. Finally, the largest feeder of the 100 was also found to have unexpectedly inaccurate forecasts (in relation to the power law fit). This feeder was found to be particularly unusual in that it supplied purely the landlord lighting for an office block. This feeder is labelled "LL Feeder" in the figure. Finally there were two other feeders with anomalous errors (labelled "Anomalous Feeders") but with poor connectivity data and hence we cannot explain the large errors. It is known that one of the feeders has a specific connection for a medical condition which could cause perhaps more irregular demands but this cannot be proven. However, these results do exhibit the important difference between true LV demand and the simple aggregation of smart meter data. In particular, different types of loads and tariffs can have a significant effect on the forecast accuracy.

A power law curve was fitted to the 88 non-OSH/non-anomalous feeders and is included in Figure 4. If the customers producing the aggregated demand were independent and identically distributed (IID) we would expect an exponent of -0.5, however we found an exponent equal to -0.47 indicating the IID assumption is not completely accurate. Further we found the variation of the customers demand also followed a power law curve very similar to the mean errors (not

shown).

5.4. Temperature Effect Analysis

Weather, in particular those related to temperature, often plays an important role in the accuracy of the load forecasts for high voltage level substations [28, 38]. In this section we consider in more detail the impact of including temperature in the forecasts. In particular, we consider both ex-ante and ex-post forecasts by utilising either forecast or actual temperature values respectively. In reality, ex-ante are the practical way to create true forecasts since, obviously the actual temperature data will not be available ahead of time. However, we include the ex-post forecasts here as well for comparison since much of the literature is based on these forms of forecast. Before presenting the analysis, it is important to note that since the temperature forecasts are generated from the fixed time point at 7AM we cannot fully compare the accuracy of the load forecasts as a function of horizon. To do this would require forecasts starting from all time periods of the day.

Table 4 shows the MAPEs for the average 4 day ahead forecasts over the test period for selected methods including their updates using temperature data, both actual and forecast values as input. From the table it is clear that the inclusion of temperature (either actual or forecast) has minimal effect on the forecast accuracy. In fact for ARWD, ARWDY and CKD-W including the temperature is detrimental to the forecast accuracy. For the ST and SnT methods, there are inconsistent results, using the actual temperature values has little to no effect on the forecast accuracy. Using the temperature forecast values, the improvement on the demand forecast is also very small with at most a 1.7% increase in forecast accuracy.

Further, Table 5 shows the accuracy of the ex-ante forecasts as a function of day ahead horizons. Also included is the MAPEs of the temperature forecasts themselves. The ARWD and ARWDY forecasts drop in accuracy by 5.6% and 4.3% respectively whereas the ST and SnT forecasts hardly change in accuracy at all. The CKD-W forecast actually improves at the 3-day ahead horizon

	Temperature Type		
Method	None	Forecast	Actual
ARWD	14.65 (4.71)	16.94 (5.26)	17.21 (5.39)
ARWDY	14.64 (4.55)	15.16 (4.43)	15.17 (4.40)
ST	15.42 (5.20)	15.16 (4.92)	15.39 (5.11)
SnT	15.66 (5.13)	15.48 (4.92)	15.66 (5.07)
CKD-W	16.54 (6.61)	17.16 (7.03)	17.02 (7.02)

Table 4: MAPEs for the methods showing the effect of including temperature data (actual or forecast) for a selection of methods.

compared to the 1-day and 2-day ahead forecasts. These results are in contrast to the accuracy of the temperature forecasts themselves, which drop in accuracy by more than 80% from one day ahead to four days ahead. If the weather was a major driver for the load we would expect a much larger drop in accuracy with horizon. Further, we also considered including up to two lags of the temperature data within the forecast methods but this also had no effect on the accuracy of the forecasts. In particular, we note that the CKD-W method naturally contains lags within the model. Further investigation into the effects of temperature lags is beyond the scope of this research and will considered in a future paper.

Moreover, we investigate the temperature effect in more detail for the ARWDY forecast as a function of feeder size. We only consider day ahead forecasts since this is when the temperature forecasts are most accurate (as shown in Table 5). To test this relationship, we plotted the difference in MAPE for the ARWDY model (with and without actual temperature) versus feeder size, see Figure E.8 in Appendix E. The plot is similar when using the forecast temperature. We chose ARWDY as it is one of the best performing model. We

	MAPE			
Method	Day 1	Day 2	Day3	Day4
ARWD	16.51 (5.08)	17.26 (5.37)	17.12 (5.37)	16.89 (5.26)
ARWDY	14.75 (4.34)	15.11 (4.41)	15.31 (4.49)	15.46 (4.50)
ST	15.12 (4.91)	15.21 (4.94)	15.16 (4.94)	15.16 (4.91)
SnT	15.47 (4.92)	15.53 (4.95)	15.46 (4.93)	15.47 (4.92)
CKD-WTf	17.22 (7.11)	17.24 (7.10)	17.02 (6.96)	17.19 (6.88)
Temperature	8.98	10.57	13.46	16.47

Table 5: MAPE Scores for different day ahead horizons for a selection of methods based on utilising forecast temperature values. Also for comparison is the average MAPE for the temperature forecasts themselves.

found that, irrespective of the feeder size, including temperature (either actual or forecast) as an explanatory variable did not result in an improvement in the out-of-sample forecast accuracy. We observe, from comparing the MAPEs for each feeder, that including the temperature forecasts only improves the errors for 19 of the 100 feeders. If we use the temperature actuals in the forecasts, this only improves a total of 20 feeders forecasts (18 of which are common to the 19 improved feeders using the forecast temperature). In all cases the MAPEs do not improve by any more than 4%.

To further test the effect of including the temperature we also consider comparing the change in the distribution of the MAPEs. We do this by using the two-sample Kolmogorov-Smirnov test (implemented using kstest2 in Matlab) at the 5% significance level. Since we have observed that the different hours of the day have different distributions of demand we split the errors according to both hour (1 to 24) and feeder. This gives us 2400 distributions to compare. Com-

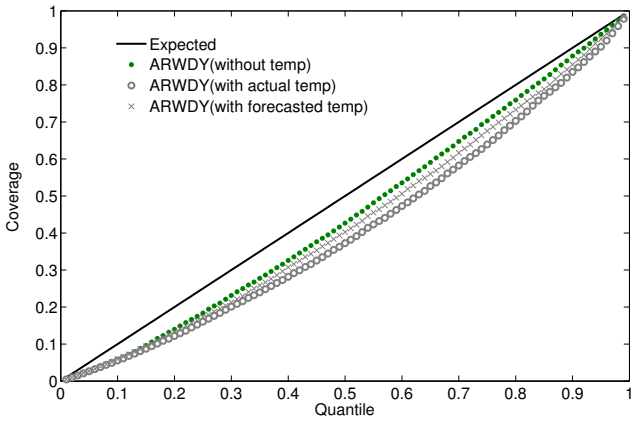


Figure 5: Reliability diagram (coverage) for the ARWDY model: (a) without temperature, (b) with actual temperature, and (c) with forecasted temperature. The solid line along the diagonal represents the expect coverage. The coverage values were average across all 100 feeders, and across all 96 horizons.

paring the ARWDY without and with the temperature forecasts as input we find that the null hypothesis (of the errors coming from the same distribution) is only rejected for a total of 79 distributions. Of these, they are split across 36 of the feeders with no more than 5 distributions failing the null hypothesis on any single feeder. Further to this, of the 79 distributions failing the null hypothesis all occur between 11PM and 6AM. In other words, when utilising temperature in the forecasts, all of the significant changes occur during the early morning hours when demand is usually more stable. There is no obvious feeder types for which the overnight errors are larger/smaller than average (and this includes to relationship to the OSH feeders in particular). In addition, when the distributions are significantly changed by including the weather, the accuracy is not improved for any of the forecasts error distributions when compared with not including temperature. Similar results hold when using the actual temperature but only 82 distributions are significantly changed (i.e. reject the null hypothesis in the ks-test).

To further access the quality of density forecasts (*coverage*), we generated

a reliability diagram, Figure 5, using the ARWDY model (with and without temperature). Overall, we found that the inclusion of temperature did not result in a reduction in model *bias* (difference between expected and obtained coverage) across all quantiles considered in this study, these results are consistent with our findings obtained using the MAPE and CRPS.

As shown in Section 3, for some feeders there is a strong correlation between the load and the temperature. However, when we include the effect in our model the accuracy either changes only slightly or, in the case of ARWD, gets worse. A major difference between the ARWD methods and ARWDY, is the lack of a seasonal term. In fact, we find that demand is much more strongly correlated with seasonality than temperature. If seasonality is a stronger driver of demand, then temperature could result in the detrimental performance in ARWD when included in the model. The strong relationship between demand, temperature and seasonality (represented by a simple sinusoidal curve) is illustrated in Figure A.7 in Appendix A. By treating the temperature as a surrogate for the seasonality the ARWD model may erroneously over-train on data which is not related to the load. Further, although the temperature data comes from within 16KM of the centre of Bracknell, this may not be localised enough for accurate forecasts at the feeder level and these short forecast horizons [53, 54].

The evidence thus suggests that at least in this area of the UK there is not a strong causal link between demand and temperature. Seasonality is a stronger driver of the demand. Interestingly at national level, results have shown that univariate models can outperform weather based models at short forecast horizons [54]. However, we are to be careful to extrapolate this further since this is a relatively small area of the UK and we only consider two months of Winter period. These findings underscore the importance of further research to investigate the relationship between LV feeder load and weather. Moreover, it is worth noting that we only had access to single point weather forecasts. Future studies may employ ensemble weather to accommodate uncertainty during the modelling.

6. Discussion

Short term load forecasts at the low voltage (LV) level are becoming increasingly important as electricity networks prepare for a low carbon future. Network solutions such as storage devices and energy management systems will require accurate forecasts to optimise the headroom and potential cost savings. Although there is a large amount of literature of short term load forecasting techniques there is not much investigation or results for LV level demand. Such demand is much more volatile and challenging than high voltage systems and there is still much to learn about the best methods and inputs for accurate forecasts.

In this paper we have presented several short term forecasting methods, both point and probabilistic, and tested them on 100 real LV feeders. We also compared them to a number of benchmarks, some of which are quite competitive. As a consequence of the studies we have found some interesting results. Accurate forecasts can be obtained by relatively simple methods, in particular, a simple average of the previous four weeks performs quite well. The best performing methods were those based on autoregression methods and a Holt-Winters-Taylor exponential smoothing method. However, it was found there was no single method which was the most accurate for a high proportion of feeders.

We also illustrated some important drivers for the accuracy of the forecast. Firstly, the size of the feeder was one of the biggest determining factor for the accuracy, with smaller feeders being more difficult to predict than larger feeders, confirming a known result from the literature. Secondly, the presence of (potentially high demand) technologies, and specifically overnight storage heaters, had significant implications in the forecast accuracy, modifying the existing power-law relationship between forecast accuracy and aggregation size. This is a novel result and could have important implications for making planning decisions such as making optimal investments in where to install storage. Thirdly, we found that the time period of the day was a major indicator of the accuracy and was

more important than the forecast horizon. Errors only increased by 2% from one day ahead to 4% for four day ahead forecasts.

In contrast to HV level load forecasts, temperature was not an important factor in the accuracy of our forecasts. We presented some detailed analysis of the results to show that the temperature either had little or no effect on the forecast accuracy, but in many cases was actually detrimental to the accuracy. We provided one potential explanation which highlighted how the strong correlation between load and temperature may in fact be describing a strong seasonal correlation which has more influence on the load than temperature.

Finally, we performed some empirical comparison of the forecasts using a variety of error measures. It was found that there was a strong correlation between the scores. Hence, the point-wise scores on the point-wise version of the forecasts could be performed to provide an accurate indication of the accuracy of the corresponding probabilistic forecast. This supports work recently in [44] and could have implications for reducing the cost of model selection for probabilistic load forecasts.

The research presented here can enhance our understanding of the low voltage network and help network operators make better management and planning decisions for future smart grids. Further the short term load forecasts results lay a foundation for methodologies and metrics to be used for generating and assessing probabilistic LV level load forecasts. In particular, there is still further understanding into the effects of weather and LCTs, the role of LV network connectivity and better probabilistic forecasts.

Acknowledgements

We would like to acknowledge the support of Scottish and Southern Electricity Networks (SSEN) and its personnel during the collaboration of the New Thames Valley Vision Project (SSET203 New Thames Valley Vision), funded by the Ofgem. We would also like to thank Dr Tamsin Lee for early discussions on the paper.

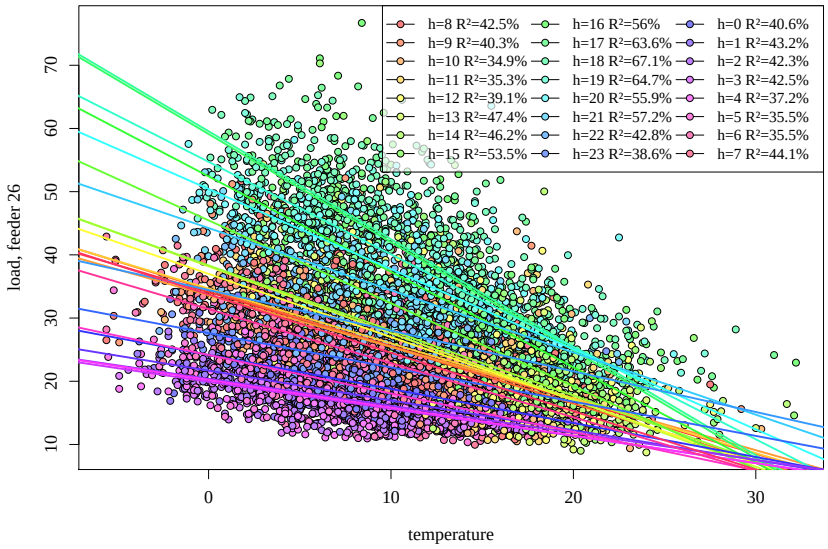


Figure A.6: Load and corresponding temperature for feeder 26. Linear fits and R^2 values are shown and colour coded according to different hourly periods of the day.

Appendix A. The Relationship between Temperature and Load

In this section further plots and information is given concerning the relationship between load and temperature. For many feeders there is a strong correlation relationship between the demand and temperature. An example of this is shown in Figure A.6 for a particular feeder in our trial. The values are broken down according to hourly periods of the day and include a basic linear fit and R^2 values. Some periods of the day have steeper gradients with respect to changes in temperature corresponding to larger increases in loads to relatively small decreases in temperature.

The temperature demand is very strongly related to annual seasonality. An example of a strong relationship between demand, temperature and seasonality (represented by a simple sinusoidal curve) is illustrated in Figure A.7. The strong correlation between seasonality and load, and between seasonality and temperature is one potential explanation for the detrimental temperature effect on the load forecast accuracy, as shown in Section 5.4. The models for load

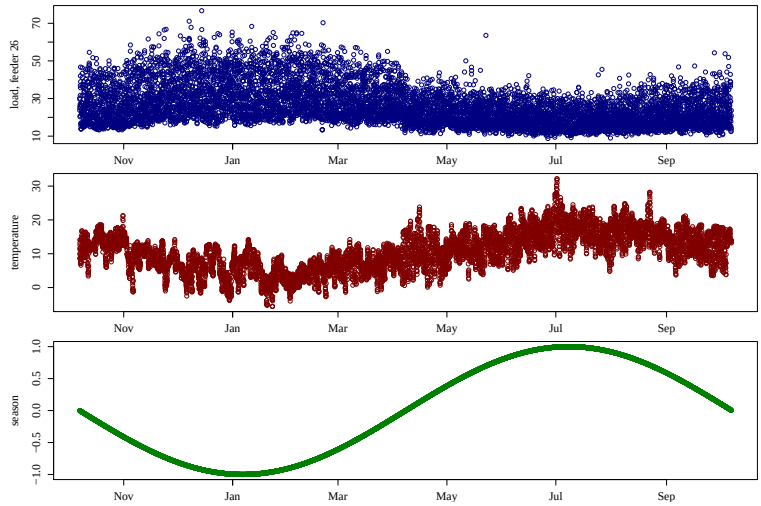


Figure A.7: Load curve (top) for feeder 26 with yearly temperature profile (centre) and a simple seasonal profile (bottom).

could be training on the seasonal effect of the temperature even if there is little to no causal link between them.

Appendix B. Details on the KDE methods

In this section we present further details of the parameters and training as used to implement the KDE methods. There are a number of available choices for the kernels used in the KDE methods. Here, we simply use the Gaussian kernel, which have been successfully used in other implementations of load forecasting [49], and secondly, evidence in the literature suggests that the choice of kernel has minimal effect on the accuracy of the forecasts [55].

The first method uses a kernel density estimate (KDE) to estimate the probability density function for each period $w = t \mod 24$ of the day, defined by the function

$$f(\hat{L}_t) = \frac{1}{(t_2 - t_1 + 1)h_L} \sum_{i \in I_w} w_i K\left(\frac{L_t - L_i}{h_L}\right), \tag{B.1}$$

where $I_w = \{i \in [t_1, t_2] | i \mod 7H = w\}$ is the index set denoting time periods

from the training set from the same period of the week as the time period t, and w_i is the weighting of historical observations i. We consider two types of KDE forecasts. The first type, denoted **KDE-W**, considers all observations having equal weight, $w_i = 1$ for all i. This forecast has one parameter, the bandwidth, h_L, for the load kernel that requires optimisation. The second type, referred to as **KDE-W**λ, favours observations around the same period of the year as time t, i.e. $w_i = \frac{\lambda^{\alpha(i)}}{\sum_{i \in I_w} \lambda^{\alpha(i)}}$, with the decay exponent $\alpha(i)$ defined by

$$\alpha(i) \quad = \quad \min\left(|\mathcal{W}(t) - \mathcal{W}(i)|, 52 - |\mathcal{W}(t) - \mathcal{W}(i)|\right), \tag{B.2}$$

where $\mathcal{W}(i) \in \{1, 2, \ldots, 52\}$ is the week of the year corresponding to the load data, L_i. Note, the first Monday of the year is defined as $\mathcal{W}(i) = 1$. Equation (B.2) is simply a periodic absolute value function with annual period, whose minimum values occur annually on the same week as the estimated day. The exponent is more relevant when there are several years of historical data. KDE-Wλ has two parameters to optimise, h_L, the bandwidth for the load kernel and $\lambda \in (0, 1]$.

Additionally, we consider kernel density estimate forecasts conditioned on independent variables y, z (CKD), such as the week-period, or weather variables e.g. the temperature or both. This is represented as

$$f(\hat{L}_t|y, z) = \sum_{i \in [t_1, t_2]} \frac{K((y_i - y)/h_y)K((z_i - z)/h_z)}{\sum_{i=1}^{n} K((y_i - y)/h_y)K((z_i - z)/h_z)} K\left(\frac{L_t - L_i}{h_L}\right) \tag{B.3}$$

where h_y, h_z are the bandwidths of the independent variables y, z respectively. If there is only one independent variable y, then one ignores the kernel of z in (B.3). CKD methods consider the whole time-series of historical observations at all time-periods.

We produce three CKD forecasts, one conditioned on the week period (**CKD-W**), a second conditioned on the week period and the actual temperature readings (**CKD-WTa**), and a third forecast conditioned on the both the week period and the forecast temperature (**CKD-WTf**). For CKD-W, $y_i = i \mod 7H$ is the week period of time interval i. CKD-W weighs observations towards similar times of the week as the forecast time-period.

The CKD methods have not been implemented with a decay parameter due to the increased computational cost. Any extra parameter increases the dimension of the parameter space and hence the computational cost for optimisation. The bandwidths and/or λ parameters of each method are found via validation. The validation period is selected to be two weeks prior to the test-period. For the KDE forecasts, all available observations before the two-week validation period are considered for training. As CKD methods are computationally more expensive, we restricted the training period to a year before the validation period. Finding the optimal parameters is a non-linear optimisation problem. For KDE-W, the *fminbnd*[3] MATLAB's built-in optimisation algorithm was used. For KDE-Wλ and all the CKD forecasts, the *fminsearchbnd*[4] optimisation package was used instead.

The load and temperature variables are normalised to $[0, 1]$ for each feeder, to accelerate the optimisation procedure. When the optimisation is complete and a forecast is produced, the normalised forecast is rescaled. To assess the effect of normalisation in the forecasts, we also experimented with KDE-W and KDE-Wλ forecasts without normalising the load. We compared the two forecasts and their errors. The forecasts were almost identical, and the error differences were marginal and evenly distributed around zero. However, the optimisation with normalised variables requires, on average, 6 less iterations than the optimisation with actual readings. For this reason, we decided to use normalised load and temperature variables for the parameter optimisation of all methods.

Appendix C. Details for the Simple Seasonal Model

In this section some extra details of the simple seasonal model described in section 4.2 are described. The dummy variables for the simple seasonal model

[3] `https://uk.mathworks.com/help/matlab/ref/fminbnd.html`

[4] `https://uk.mathworks.com/matlabcentral/fileexchange/8277-fminsearchbnd--fminsearchcon`

representing the daily and weekly hourly periods are defined by

$$\mathcal{D}_j(t) = \begin{cases} 1, & t \mod H = j, \\ 0, & \text{otherwise}, \end{cases}$$

and

$$\mathcal{W}_j(t) = \begin{cases} 1, & t \mod 7H = j, \\ 0, & \text{otherwise}, \end{cases}$$

respectively. For this model there are essentially 168 models representing each period of the week each with average, linear trend, and annual seasonality terms. The a_k terms represent the average demand for that hourly period (which is augmented based on the day of the week by f_l), a linear trend term b_k, and annual seasonality terms defined by c_k and d_k. The parameters for each hour and each quantile are found by a quantile regression over the historical data using the pinball function [56].

Appendix D. Details for the ARWD and ARWDY Methods

There are two mean models used to define the residual time series described in section 4.3. The first, ARWD, estimates a simple weekly average and can be written

$$\mu_t = \sum_{j=1}^{7H} \beta_j \mathcal{W}_j(t). \tag{D.1}$$

where $\mathcal{W}_j(t)$ is the period of the week dummy variable as defined in Appendix Appendix C. The weekly mean parameters are β_j and hence the mean equation (D.1), μ_k, is estimated by ordinary least squares (OLS) over the initial prior year of historical loads.

The second model, ARWDY, updates the simple weekly average to include an annual seasonality.

$$\mu_t = \sum_{j=1}^{7H} \beta_j \mathcal{W}_j(t) + \sum_{k=1}^{K} \alpha_{1,k} \sin(2\pi tk/A) + \alpha_{2,k} \cos(2\pi tk/A) \tag{D.2}$$

with parameters β_j and $\alpha_{j,k}$, and $A = 365H$ as annual seasonality. So the annual seasonality is additionally modelled by a Fourier approximation of order

K (we choose only $K = 2$). The dummy variable $\mathcal{W}_j(t)$ is as before. As with the ARWD model the μ_k is estimated by OLS linear regression fit to the training data.

The methods can be updated to generate probabilistic forecast methods. The methods are not quite as straightforward as the point forecasts due to a two-step process. We present a related but alternative autoregressive probabilistic forecast technique for the ARWDY case (which generalises trivially for the ARWD case).

The following method creates confidence by modelling the variance of the model from the residuals. As before an autoregression is performed on the residual equation (2) but now the mean load is modelled by the following equation.

$$\mu_t = \sum_{j=1}^{7H} \beta_j \mathbb{W}_j(t) + \sum_{k=1}^{K} \gamma_k \mathbb{D}_j(t) + \alpha_{1,k} \mathbb{D}_j(t) \sin(2\pi tk/A) + \alpha_{2,k} \mathbb{D}_j(t) \cos(2\pi tk/A), \tag{D.3}$$

where $A = 365H$. The annual seasonality is modelled by a Fourier approximation of order $K = 2$. This time a different (but equivalent) basis function is used with dummy variables for different horizon periods modelled as

$$\mathbb{W}_j(t) = \begin{cases} 1, & \text{t mod 7H} \leq j \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad \mathbb{D}_j(t) = \begin{cases} 1, & \text{t mod H} \leq j \\ 0, & \text{otherwise} \end{cases}.$$

The mean equation (D.3) is solved using a lasso method tuned to ensure that the Hannan-Quinn information criterion (HQC) is minimised. The past year is used in the training. The conditional mean equation (D.1) is then estimated using the discovered residuals by solving the Yule-Walker equations. As before, the order is chosen by minimising the Akaike information criterion. To create the confidence bounds we model the error term (the residual from equation (D.3)) as

$$\epsilon_t = \sigma_t Z_t, \tag{D.4}$$

where σ_t is the conditional standard deviation of ϵ_t and $(Z_t)_{t_\in \mathbb{Z}}$ is an iid random variable with $\mathbb{E}(Z_t) = 0$ with $\mathbb{V}ar(Z_t) = 1$. We then assume a similar relation-

ship for different horizon periods in the model for the standard deviation as we do for the mean,

$$\sigma_t = \sum_{j=1}^{7H} \tilde{\beta}_j \mathbb{W}_j(t) + \sum_{k=1}^{K} \tilde{\gamma}_k \mathbb{D}_j(t) + \tilde{\alpha}_{1,k} \mathbb{D}_j(t) \sin(2\pi tk/A) + \tilde{\alpha}_{2,k} \mathbb{D}_j(t) \cos(2\pi tk/A), \tag{D.5}$$

Although we do not know the values of the variance we can model a scaled version of σ_t by using the model in equation (D.5) to fit via the HQC to minimise the lasso problem for $|\epsilon_t|$. Since $\mathbb{E}(|\epsilon_t|) = \sigma_t \mathbb{E}(|Z_t|))$ this means we are actually estimating $C\sigma_t$ for some constant C. We estimate the constant by considering the residuals $\epsilon_t / C\sigma_t$ which, given (D.4), should therefore behave like a scaled version of Z_t. Considering the variance of these residuals and noting that Z_t has variance one we use this to estimate the constant and hence find σ_t. We now calculate the standardised residuals ϵ_t/σ_t to put an empirical distribution on Z_t and estimate the quantiles (which once scaled by σ_t give the quantiles on ϵ_t). Note, if Z_t follows a standard normal distribution then C would be $\sqrt{(2/pi)} = 0.798$. For our data we found this was usually smaller, between 0.5 and 0.75. An advantage of this method is the quick computational speed. Computing 99-percentiles for 4-days ahead takes about 2 seconds per feeder.

Appendix E. Further Temperature Results

Temperature data is a key variable for inclusion within short term load forecasts at higher voltage and system level [18], [19]. In contrast, the effect at the low voltage level is not fully understood, in particular, whether there is a weather effect for different size feeders. Figure E.8 shows the difference between the average MAPE errors when using and not-using the actual temperature within the ARWDY forecast methods for each of the 100 feeders in the test set. Also included is a power law fit. There does not appear to be a strong relationship between the improvement by including temperature data within the forecast and the size of the feeder in this case study. A similar result holds when we utilise weather forecast data rather than actuals.

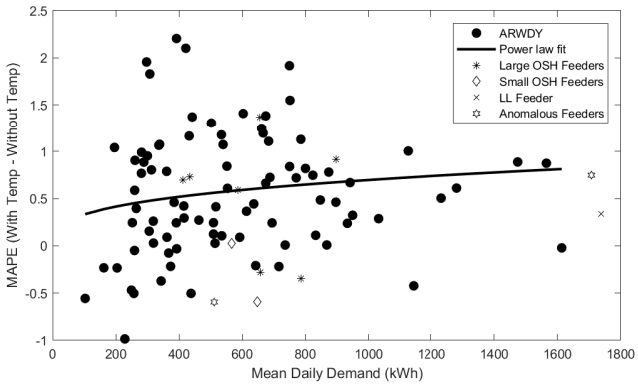


Figure E.8: Difference in MAPE for the ARWDY model (with and without actual temperature) plotted against feeder size. The anomalous feeders found in section 5.3 have been labeled separately.

[1] Y. Liu, W. Wang, N. Ghadimi, Electricity load forecasting by an improved forecast engine for building level consumers, Energy 139 (2017) 18–30–3052.

[2] M. Rowe, T. Yunusov, S. Haben, W. Holderbaum, B. Potter, The real-time optimisation of DNO owned storage devices on the LV network for peak reduction, Energies 7 (6) (2014) 3537–3560. doi:10.3390/en7063537.

[3] M. Rowe, T. Yunusov, S. Haben, C. Singleton, W. Holderbaum, B. Potter, A peak reduction scheduling algorithm for storage devices on the low voltage network, IEEE Trans. on Smart Grid 5 (2014) 2115–2124.

[4] W. El-Baz, P. Tzscheutschler, Short-term smart learning electrical load prediction algorithm for home energy management systems, Applied Energy 147 (2015) 10 – 19.

[5] H. Chitsaz, H. Shaker, H. Zareipour, D. Wood, N. Amjady, Short-term electricity load forecasting of buildings in microgrids, Energy and Buildings 99 (Supplement C) (2015) 50 – 60.

[6] C. Bennett, R. A. Stewart, J. Lu, Autoregressive with exogenous variables and neural network short-term load forecast models for residential low voltage distribution networks, Energies 7 (5) (2014) 2938–2960.

[7] G. Giasemidis, S. Haben, T. Lee, C. Singleton, P. Grindrod, A genetic algorithm approach for modelling low voltage network demands, Applied Energy 203 (2017) 463–473.

[8] S. B. Taieb, J. W. Taylor, R. J. Hyndman, Hierarchical probabilistic forecasting of electricity demand with smart meter data (2017) 1–30.
URL https://robjhyndman.com/papers/HPFelectricity.pdf

[9] T. K. Wijaya, M. Vasirani, S. Humeau, K. Aberer, Cluster-based aggregate forecasting for residential electricity demand using smart meter data, in: 2015 IEEE International Conference on Big Data (Big Data), 2015, pp. 879–887.

[10] X. Sun, P. B. Luh, K. W. Cheung, W. Guan, L. D. Michel, S. S. Venkata, M. T. Miller, An efficient approach to short-term load forecasting at the distribution level, IEEE Transactions on Power Systems 31 (4) (2016) 2526–2537.

[11] L. Hattam, D. V. Greetham, Green neighbourhoods in low voltage networks: measuring impact of electric vehicles and photovoltaics on load profiles, Journal of Modern Power Systems and Clean Energy 5 (1) (2017) 105–116. doi:http://doi.org/10.1007/s40565-016-0253-0.
URL http://dx.doi.org/10.1007/s40565-016-0253-0

[12] R. Sevlian, R. Rajagopal, Short Term Electricity Load Forecasting on Varying Levels of Aggregation, ArXiv e-prints.

[13] H. Alfares, M. Nazeeruddin, Electric load forecasting: literature survey and classification of methods, International Journal of Systems Science 33 (2002) 23–34.

[14] J. W. Taylor, A. Espasa, Energy forecasting, International Journal of Forecasting 24 (2008) 561565.

[15] T. Hong, P. Pinson, S. Fan, Global energy forecasting competition 2012, International Journal of Forecasting 30 (2) (2014) 357 – 363. doi:https://doi.org/10.1016/j.ijforecast.2013.07.001.
URL http://www.sciencedirect.com/science/article/pii/S0169207013000745

[16] P. Gaillard, Y. Goude, R. Nedellec, Additive models and robust aggregation for gefcom2014 probabilistic electric load and electricity price forecasting, International Journal of Forecasting 32 (3) (2016) 1038 – 1050. doi:https://doi.org/10.1016/j.ijforecast.2015.12.001.
URL http://www.sciencedirect.com/science/article/pii/S0169207015001545

[17] P. Wang, B. Liu, T. Hong, Electric load forecasting with recency effect: A big data approach, International Journal of Forecasting 32 (3) (2016) 585 – 597. doi:https://doi.org/10.1016/j.ijforecast.2015.09.006.
URL http://www.sciencedirect.com/science/article/pii/S0169207015001557

[18] M. Bessec, J. Fouquau, The non-linear link between electricity consumption and temperature in europe: A threshold panel approach, Energy Economics 30 (5) (2008) 2705 – 2721.

[19] M. Dahl, A. Brun, G. B. Andresen, Using ensemble weather predictions in district heating operation and load forecasting, Applied Energy 193 (2017) 455 – 465.

[20] F. L. Quilumba, W. J. Lee, H. Huang, D. Y. Wang, R. L. Szabados, Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities, IEEE Transactions on Smart Grid 6 (2) (2015) 911–918. doi:10.1109/TSG.2014.2364233.

[21] V. Dehalwar, A. Kalam, M. L. Kolhe, A. Zayegh, Electricity load forecasting for urban area using weather forecast information, in: 2016 IEEE International Conference on Power and Renewable Energy (ICPRE), 2016, pp. 355–359.

[22] Irish Social Science Data Archive, Cer smart metering project (2012).
URL http://www.ucd.ie/issda/

[23] S. Haben, J. A. Ward, D. V. Greetham, P. Grindrod, C. Singleton, A new error measure for forecasts of household-level, high resolution electrical energy consumption, Int. J. of Forecasting 30 (2014) 246–256.

[24] J. M. C. Sousa, L. M. P. Neves, H. M. M. Jorge, Short-term load forecasting using information obtained from low voltage load profiles, in: 2009 International Conference on Power Engineering, Energy and Electrical Drives, 2009, pp. 655–660. doi:10.1109/POWERENG.2009.4915229.

[25] A. Veit, C. Goebel, R. Tidke, C. Doblander, H.-A. Jacobsen, Household electricity demand forecasting: Benchmarking state-of-the-art methods, in: Proceedings of the 5th International Conference on Future Energy Systems, e-Energy '14, ACM, New York, NY, USA, 2014, pp. 233–234.

[26] M. Ghofrani, M. Hassanzadeh, M. Etezadi-Amoli, M. S. Fadali, Smart meter based short-term load forecasting for residential customers, in: IEEE North American Power Symposium (NAPS), Boston, USA, 2011, pp. 1–5.

[27] C. N. Yu, P. Mirowski, T. K. Ho, A sparse coding approach to household electricity demand forecasting in smart grids, IEEE Transactions on Smart Grid 8 (2) (2017) 738–748.

[28] J. D. Hobby, G. H. Tucci, Analysis of the residential, commercial and industrial electricity consumption, in: Innovative Smart Grid Technologies Asia (ISGT), 2011 IEEE PES, Perth, Australia, 2011.

[29] C. J. Bennett, R. A. Stewart, J. W. Lu, Forecasting low voltage distribution network demand profiles using a pattern recognition based expert system, Energy 67 (Supplement C) (2014) 200 – 212.

[30] N. Ding, C. Benoit, G. Foggia, Y. Bsanger, F. Wurtz, Neural network-based model design for short-term load forecast in distribution systems, IEEE Transactions on Power Systems 31 (1) (2016) 72–81.

[31] D. Alberg, M. Last, Short-Term Load Forecasting in Smart Meters with Sliding Window-Based ARIMA Algorithms, Springer International Publishing, Cham, 2017, pp. 299–307.

[32] S. Humeau, T. K. Wijaya, M. Vasirani, K. Aberer, Electricity load forecasting for residential customers: Exploiting aggregation and correlation between households, in: 2013 Sustainable Internet and ICT for Sustainability (SustainIT), 2013, pp. 1–6. doi:10.1109/SustainIT.2013.6685208.

[33] T. Zufferey, A. Ulbig, S. Koch, G. Hug, Forecasting of Smart Meter Time Series Based on Neural Networks, Springer International Publishing, Cham, 2017, pp. 10–21.

[34] C. Alzate, M. Sinn, Improved electricity load forecasting via kernel spectral clustering of smart meters, in: 2013 IEEE 13th International Conference on Data Mining, 2013, pp. 943–948. doi:10.1109/ICDM.2013.144.

[35] B. Hayes, J. Gruber, M. Prodanovic, Short-term load forecasting at the local level using smart meter data, in: 2015 IEEE Eindhoven PowerTech, 2015, pp. 1–6.

[36] O. Valgaev, F. Kupzog, H. Schmeck, Low-voltage power demand forecasting using k-nearest neighbors approach, in: 2016 IEEE Innovative Smart Grid Technologies - Asia (ISGT-Asia), 2016, pp. 1019–1024. doi:10.1109/ISGT-Asia.2016.7796525.

[37] T.-H. Dang-Ha, F. M. Bianchi, R. Olsson, Local Short Term Electricity Load Forecasting: Automatic Approaches, ArXiv e-prints.

[38] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, R. J. Hyndman, Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond, International Journal of Forecasting 32 (2016) 896–913.

[39] T. Hong, S. Fan, Probabilistic electric load forecasting: A tutorial review, International Journal of Forecasting 32 (3) (2016) 914 – 938.

[40] S. Arora, J. Taylor, Forecasting electricity smart meter data using conditional kernel density estimation, Omega 59 (2016) 47–59.

[41] A. Gerossier, R. Girard, G. Kariniotakis, A. Michiorri, Probabilistic day-ahead forecasting of household electricity demand, in: Proceedings of the 24th International Conference on Electricity Distribution (CIRED), 2017.

[42] S. B. Taieb, R. Huser, R. J. Hyndman, M. G. Genton, Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression, IEEE Transactions on Smart Grid 7 (5) (2016) 2448–2455.

[43] Z. Li, A. Hurn, A. Clements, Forecasting quantiles of day-ahead electricity load, Energy Economics 67 (2017) 60 – 71.

[44] J. Xie, T. Hong, Comparing two model selection frameworks for probabilistic load forecasting, in: 2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), 2016, pp. 1–5.

[45] S. Rahman, Formulation and analysis of a rule-based short-term load forecasting algorithm, in: Proceedings of the IEEE, Vol. 78, 1990, pp. 805–816.

[46] T. Hong, Long term probabilistic load forecasting and normalization with hourly information, IEEE Transactions on Smart Grid 5 (2014) 456–462.

[47] F. Ziel, R. Steinert, S. Husmann, Efficient modeling and forecasting of electricity spot prices, Energy Economics 47 (2015) 98 – 111. doi:https://doi.org/10.1016/j.eneco.2014.10.012.
URL http://www.sciencedirect.com/science/article/pii/S0140988314002576

[48] T. Yunusov, S. Haben, T. Lee, F. Ziel, W. Holderbaum, B. Potter, Evaluating the effectiveness of storage control in reducing peak demand on low voltage feeders, in: Proceedings of the 24th International Conference on Electricity Distribution (CIRED), 2017.

[49] S. Haben, G. Giasemidis, A hybrid model of kernel density estimation and quantile regression for gefcom2014 probabilistic load forecasting, Int. J. Forecasting 32 (2016) 1017–1022.

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

[51] S. Arora, J. Taylor, Rule-based autoregressive moving average models for forecasting load on special days: A case study for france, European Journal of Operational Research 266 (2017) 259–268.

[52] T. Gneiting, A. E. Raftery, Strictly proper scoring rules, prediction, and estimation, Journal of the American Statistical Association 102 (2007) 359–378.

[53] T. Hong, P. Wang, L. White, Weather station selection for electric load forecasting, International Journal of Forecasting 31 (2) (2015) 286 – 295.

[54] J. W. Taylor, An evaluation of methods for very short term electricity demand forecasting using minute-by-minute british data, International Journal of Forecasting 24 (2008) 645–658.

[55] J. Jeon, J. W. Taylor, Using conditional kernel density estimation for wind power density forecasting, Journal of the American Statistical Association 107 (497) (2012) 66–79.

[56] R. Koenker, G. B. Jr., Regression quantiles, Econometrica 46 (1978) 33–50.