A Probabilistic Framework for Forecasting Household Energy Demand Profiles

N. Charlton *[1], S. A. Haben †[2], P. Grindrod[2], D. V. Greetham[3], and C. Singleton[1]

[1]CountingLab Ltd, Reading
[2]Mathematical Institute, University of Oxford
[3]Department of Mathematics and Statistics, University of Reading

March 10, 2014

Abstract

Recent work introduced a novel time-permuting error measure for forecasts of half-hourly, household-level energy demand, designed to reward forecasts which predict extremes (spikes) in demand at approximately the right times, albeit perhaps slightly early or late. In many applications such as smart storage control, such forecasts are preferable to those that predict no spikes at all. Building on that idea, we make three contributions. First we introduce a probabilistic framework to estimate error distributions for actuals about forecasts, using the time-permuting error measure. The framework includes a variable discount for older, possibly less relevant data. Second we employ this framework to derive conditions to be satisfied by the optimal forecast under the time-permuting error measure. In turn this requires a mixture of discrete (non derivative) optimisation and calculus to condition forecasts on available historical observations. Finally we demonstrate the usefulness of our framework by using it to forecast the daily energy demand profiles for a large number of domestic energy customers. In particular we illustrate how such customers might be classified according to the relative forecastability of their behaviour and the corresponding need for different amounts of history to achieve such forecasts.

*n.a.charlton@reading.ac.uk
†haben@maths.ox.ac.uk

1 Introduction

In previous work (Haben et al., 2014) the authors introduced a novel error measure that favours the forecasting of extremes (spikes) in behavioural demand profiles, by allowing for a relatively limited amount of time shifting, permuting the forecast to match the actual (possibly under some penalty). This is designed to reward methods that predict extreme sharp spikes in demand yet may be slightly imprecise in the timing of such events. For example, if a method estimates a sharp spike in demand but the actual profile contains such a spike but at a slightly different time, then a regular norm penalises that forecast twice, once for the forecast spike and once for the missed actual spike (Keil & Craig, 2009). Thus when such simple error norms are used it may have been *preferable* to forecast no spike at all and be penalised only for the miss. The measure introduced by Haben et al. (2014) removes that disincentive.

Here we shall consider a variation of that type of measure that biases the consequent forecaster to predict early rather than late spikes. In some applications, where warning or preparation time of such extremes is desired, a conservative warning forecast is advantageous. In other applications, such as the smart control of local energy storage and its subsequent release, the forecaster-controller should not rely on low demand periods immediately prior to predicted high demand to increase the storage demand (charging up), since an early actual spike, even if unlikely, will result in possibly violating the available supply headroom. In contrast, early-peak biased forecasts can enable sufficient planning of the storage device to successfully reduce peak demands on the network (Molderink et al., 2010; Rowe et al., 2012; Haben et al., 2013).

The goal of this paper is threefold. First we introduce a probabilistic framework to estimate error distributions for actuals about forecasts that includes a variable discount for older, possibly less relevant data, and incorporates the error measures (and thus the need for discrete permutation matching) given in Haben et al. (2014). Second, we employ this framework to derive conditions to be satisfied by the optimal forecast under those defined penalties. In turn this requires a mixture of discrete (non derivative) optimisation and calculus to condition forecasts on available historical observations. We may also set a prior forecast to provide a relatively insensitive start-up when there is little or no history. Finally, we shall demonstrate the usefulness of that method by forecasting the daily energy demand profiles for a large number of domestic energy customers using their historical half hourly electricity usage. In particular we shall illustrate how such customers might be classified according to the relative forecastability of their behaviour and the corresponding need for different amounts of history to achieve such forecasts.

This method may be useful in a wide variety of applications where the profile data to be forecast is spikey and volatile and where early forecasting of such spikes is desirable. In considering the example application to household energy demand an efficient (low cost) and accurate (sometimes conservative) forecast method allowing for rapid calibration (with as little history as possible), for both individual users (households), or groups of users (at substations), is highly desirable on both sides of the electrical meter. Such a method is urgently required as novel, innovative, smart low carbon technologies are introduced by both consumers and by distribution network operators (Haben et al., 2013).

2 A probabilistic framework

In this section we introduce a Bayesian framework for deriving forecasts based on minimising a generalised version of the error measure introduced in previous work (Haben et al., 2014). Suppose that we wish to forecast a demand profile for a single *day* (or some other time period), where each forecast, and each observation, comprises of a vector (time series) of $m > 1$ successive point measurements (for example, one measurement for each of the $m = 48$ half hours per day). Then we may make a direct comparison between any forecast vector and the actual observed profile for that *day*.

Consider the error measure introduced in Haben et al. (2014). This is defined so as to minimise the penalty of forecasting peaks within the day if such peaks are only slightly mistimed (by say a couple of hours) compared to the actual peak. It allows for elements of the forecast vector to be permuted so as to better fit the observed profile. Here we refine that model to include a cost, weighted by the point demand for the time shifting of each point.

Let $\mathbf{f} = (f_1, ..., f_m)^T$ and $\mathbf{a} = (a_1, ..., a_m)^T$ denote a forecast and an actual observed profile respectively. For standard domestic smart meters typically $m = 48$ so readings are every half hour (for instance, the planned smart meter rollout in the United Kingdom will provide data at that resolution (Department of Energy and Climate Change (DECC), 2013)). However, for higher resolution monitoring in other applications we may set m much larger. Then we define

$$E(\mathbf{f}, \mathbf{a}) = \min_S \left(\sum_{i=1}^{m} |f_{S^{-1}(i)} - a_i|^p + a_i^p g(S^{-1}(i) - i) \right)^{1/p} \quad (1)$$

to be the error measure. Here S ranges over all suitable permutations of the index set $\{1, 2, ..., m\}$. The error measure (1) is a combination of the difference between the magnitudes of the permuted forecast and the actual (the first term), and a penalty for the displacement of the forecast itself (the second term). If $\mathbf{f}$ has a peak at $j = S^{-1}(i)$, which is matched by S to a corresponding peak, say a_i, at $i = S(j)$, then if $j < i$ the peak is forecast early, which can be preferable than when $j > i$, and it is forecast late. So here the shifting-penalty function, $g(m)$, must be such that

(i) $g(0) = 0$,

(ii) $g(m)$ is increasing for $m > 0$,

(iii) $g(m)$ is decreasing for $m < 0$,

(iv) $0 < g(-m) < g(m)$ for any $m > 0$

The parameter $p \geq 1$ biases the measure to penalise errors at peaks: we shall take $p = 4$ below. The calculation of $E(\mathbf{f}, \mathbf{a})$ thus requires a fast search over permutations, S, that can be achieved efficiently using the Hungarian algorithm (Munkres, 1957; Haben et al., 2014) or, for some classes of g, more efficient graph-theoretic methods (Charlton et al., 2013).

Now suppose we are given a set of K actuals (historical daily observations) given by $\{\mathbf{a}_k | k = 1, 2, ..., K\}$, and the corresponding ages for each observation, without loss of generality in increasing order, say $\{t_k | k = 1, 2, ..., K\}$.

Bayes' theorem says

$$P(\mathbf{f} | \{\mathbf{a}_k, t_k\}) \propto P(\{\mathbf{a}_k, t_k\} | \mathbf{f}).P_{prior}(\mathbf{f}). \tag{2}$$

The middle, "model", term describes the distribution of errors (in the observations) given the forecast. In general this term may contain some other parameters $(\sigma, \lambda, ...)$ (see below) which must be determined along with the forecast $\mathbf{f}$. So we can extend equation (2) to include such parameters and write

$$P((\mathbf{f}, \sigma, \lambda) | \{\mathbf{a}_k, t_k\}) \propto P(\{\mathbf{a}_k, t_k\} | (\mathbf{f}, \sigma, \lambda)).P_{prior}((\mathbf{f}, \sigma, \lambda)). \tag{3}$$

Let us consider the following negative exponential model for a single daily observation $\mathbf{a}$ of age t:

$$P(\{\mathbf{a}, t\} | (\mathbf{f}, \sigma, \lambda)) = e^{-E(\mathbf{f}, \mathbf{a}) \lambda^t / \sigma} \frac{\lambda^t}{\sigma}. \tag{4}$$

Here $\lambda \in (0, 1]$ is a discounting factor (due to age), and $\sigma > 0$ is the standard deviation (at $t = 0$). The larger the observation age, t, the larger the expected (tolerable) value (and variance) of the distribution for the errors (1) at that age. For this negative exponential distribution we have mean $= \sigma/\lambda^t \to \infty$ and variance $= \sigma^2/\lambda^{2t} \to \infty$ as $t \to \infty$.

If the observations $(\mathbf{a}_k, t_k)$ are independent, we can write

$$P(\{\mathbf{a}_k, t_k\}|(\mathbf{f}, \sigma, \lambda)) = \exp\left(-\sum_{k=1}^{K} \lambda^{t_k} E(\mathbf{f}, \mathbf{a}_k)/\sigma + \ln\lambda \sum_{k=1}^{K} t_k - K \ln\sigma\right). \tag{5}$$

Assuming for the moment that the prior in (3), $P_{prior}(\mathbf{f}, \sigma, \lambda)$, is uniform and plays no role: then we maximise the posterior distribution (5), which is equivalent to finding maximum likelihood estimates (MLEs) for the unknowns, $(\sigma, \lambda, \mathbf{f})$, that jointly maximize the model for the observations.

We therefore minimize the negative of the log-likelihood, denoted by,

$$H(\sigma, \lambda, \mathbf{f}) = \sum_{k=1}^{K} \lambda^{t_k} E(\mathbf{f}, \mathbf{a}_k)/\sigma - \ln\lambda \sum_{k=1}^{K} t_k + K \ln\sigma. \tag{6}$$

Now we differentiate H with respect to σ and find that

$$\sigma = \frac{1}{K} \sum_{k=1}^{K} \lambda^{t_k} E(\mathbf{f}, \mathbf{a}_k). \tag{7}$$

Similarly we differentiate H with respect to λ to find that

$$\sigma = \frac{\sum_{k=1}^{K} t_k \lambda^{t_k} E(\mathbf{f}, \mathbf{a}_k)}{\sum_{k=1}^{K} t_k}. \tag{8}$$

From (7) and (8) we see that λ must satisfy

$$Q(\lambda) \equiv \sum_{k=1}^{K} \lambda^{t_k} E(\mathbf{f}, \mathbf{a}_k)(t_k - \bar{t}) = 0$$

where $\bar{t} = (t_1 + ..t_K)/K$ is the mean observation time. Clearly $Q(0) = 0$ and $Q(\lambda) \sim \lambda^{t_1} E(\mathbf{f}, \mathbf{a}_1)(t_1 - \bar{t}) < 0$ for small λ. Similarly large λ we have $Q > 0$. Thus Q has at least one root for $\lambda > 0$.

Hence if $\mathbf{f}$ is given, we may solve (7) and (8) together for σ and λ. If the solution is such that $\lambda > 1$, then we must set $\lambda = 1$ (no discounting) and σ is given by (7) and in that case: $\sigma = \frac{1}{K}\sum_{k=1}^{K}\ E(\mathbf{f}, \mathbf{a}_k)$. Finally in all cases we must search for $\mathbf{f}$ so as to minimize H.

Substituting from (7) into (6) we obtain a simplified objective

$$H(\sigma, \lambda, \mathbf{f}) = -\ln\lambda \sum_{k=1}^{K} t_k + K(1 + \ln\sigma). \tag{9}$$

In deploying any discrete search algorithm (for example genetic algorithms (GA), Nelder-Mead method or any other non derivative optimization methods (Mitchell, 1998; Conn et al., 2009) we may trial any suitable $\mathbf{f}$, calculate the errors

$$E(\mathbf{f}, \mathbf{a}_k) \quad k = 1, ..., K;$$

then next solve (7) and (8) together for σ and λ, and then finally use these values in the objective function (9). If we obtain a maximum where $\lambda \to 0$ we are effectively using the most recent (least aged) observation as the forecast. If we obtain a maximum where $\lambda = 1$ then we are using all observations equally, regardless of age. For intermediate values of λ we get some idea of how much history is really relevant to a forecast. For example we could ignore observations for which $t_k > t_1 - 2/\log_{10}\lambda$, where the observations have weights less than one percent of that of the most recent observation.

The parameter σ measures the overall size of errors and is small for consistent and forecastable, or large for inconsistent and volatile households. Hence the values of σ and λ can be used to classify different demand profiles in terms of their forecastability and amount of history required to produce the forecast.

It may well seem appropriate to use uniform priors for $\lambda \in (0, 1]$ and $\sigma > 0$ but if K (the number of observations) is small we may wish to have a prior for $\mathbf{f}$, so as the number of observations increases the prior estimate exerts less influence on the forecast. The easiest way is to assume

$$P_{prior}(\mathbf{f}) = \frac{e^{-E(\mathbf{f}, \mathbf{f}^*)/\sigma_0}}{\sigma_0},$$

for some $\sigma_0 > 0$ and $\mathbf{f}^*$, a modal prior forecast, given.

Then H (from the posterior) in (9) must be augmented with an extra term and becomes

$$H(\sigma, \lambda, \mathbf{f}) = -\ln \lambda \sum_{k=1}^{K} t_k + K(1 + \ln \sigma) + E(\mathbf{f}, \mathbf{f}^*)/\sigma_0,$$

while equations (7) and (8) must still hold as before. Clearly for little or no data we will have $\mathbf{f} = \mathbf{f}^*$, the prior modal estimate. The last term dominates depending on both K and σ.

3 Example with Smart Meter Data

We apply the forecast methodology described in Section 2 to real smart meter data from the Irish smart meter trial (Irish Social Science Data Archive, 2012). To define our error measure we take $p = 4$ in the error equation (1) to weight the model to favour larger peaks and define $g : \mathbb{Z} \longrightarrow \mathbb{R}$ by

$$g(m) = \begin{cases} -0.05m & \text{if } m \leq 0 \\ 0.1m & \text{if } m > 0 \end{cases} \tag{10}$$

Thus we penalise late forecast peaks more than early forecast peaks.

We forecast each day of a single week for 543 domestic customers[1] using the same day of the previous 15 weeks giving us 3801 forecasts in total. We use a uniform prior and approximate the maximum likelihood estimates $(\sigma, \lambda, \mathbf{f})$ using a genetic algorithm.

To test that the errors $E(\mathbf{f}, \mathbf{a}_0)$ at time $t_0 = 0$ are distributed according to the exponential model (4) we consider the 95^{th} percentile of the errors, $\hat{E} = \sigma \ln(20)$. We found that 120 of the 3801 errors in the forecasts exceed the 95% quantile, corresponding to a p-value of 0.99999999.

3.1 Model Parameters

The model parameters σ and λ play an important role in determining the accuracy of the data and quantity of historical data used to create the forecast. Figure 1 shows the relationship between λ and σ for 700 household forecasts. In general less variable customers (smaller σ) tend to use fewer days of their history (have smaller λ values) in formulating the forecast. Similarly those customers which use most of their history (have λ values closer to 1) also tend to be the more variable customers (larger σ).

3.2 Examples

The forecast outperforms two alternative, naive forecasting methodologies: a simple persistence forecast using the previous week as the forecast, and another using the standard mean profile from the previous 10 weeks. The new maximum likelihood, early bias forecast scored lower adjusted errors (1) for 2924 out of 3801 forecasts compared to the persistence forecast, and for 2677 out of 3801 forecasts compared to the mean forecast. Due to the lack of flexibility in the mean forecast, consistent peaks are often underestimated (see Haben et al. (2014)).

The exemplars in Figures 2-4 shows the daily profiles of the 10 historic weeks and then the final forecasted week together with the actual at the bottom. The plots show some of the success of the method in not only

[1] We used all domestic control customers for whom there is complete data from 3rd May 2010 to 3rd Oct 2010.

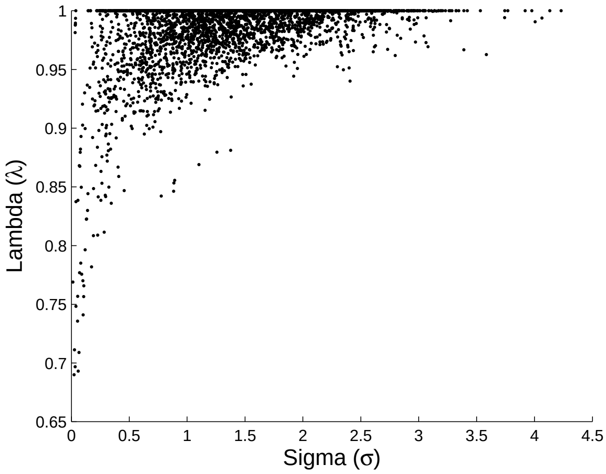

Figure 1: Relationship between σ and λ for 700 household level forecasts.

forecasting the correct magnitude of the peaks but also the correct approximate timing. In particular, the examples show the tendency of the method to favour forecasting peaks early if such peaks are present in the historical data. For example, the historical data in Figure 2 show a peak around the 20^{th} half hour (10am) but due to the appearance of a peak prior to 10am in 2 of the historical profiles the forecast has successfully anticipated an early peak for the forecasted week.

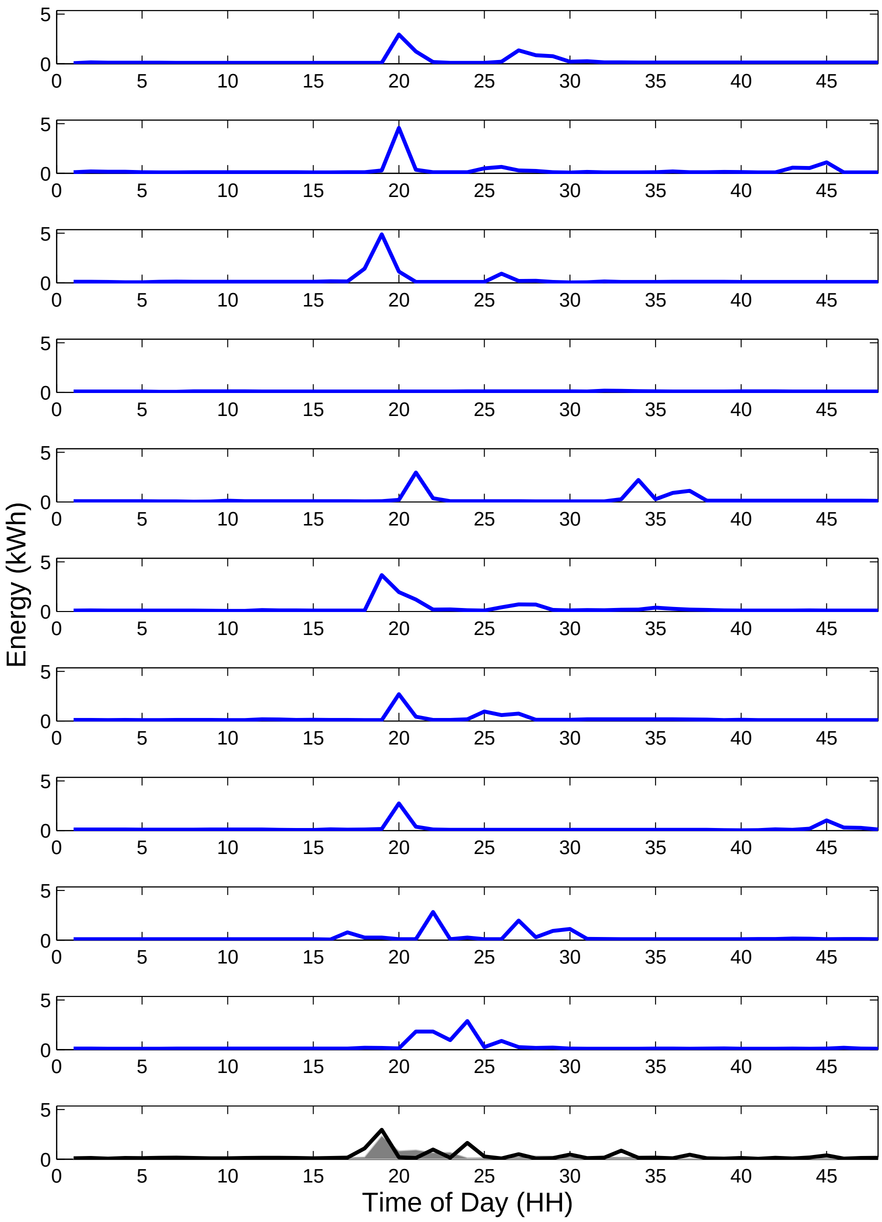

Figure 2: Forecast 202 (bold) against actual (shaded) with historical profiles (top 10 plots).

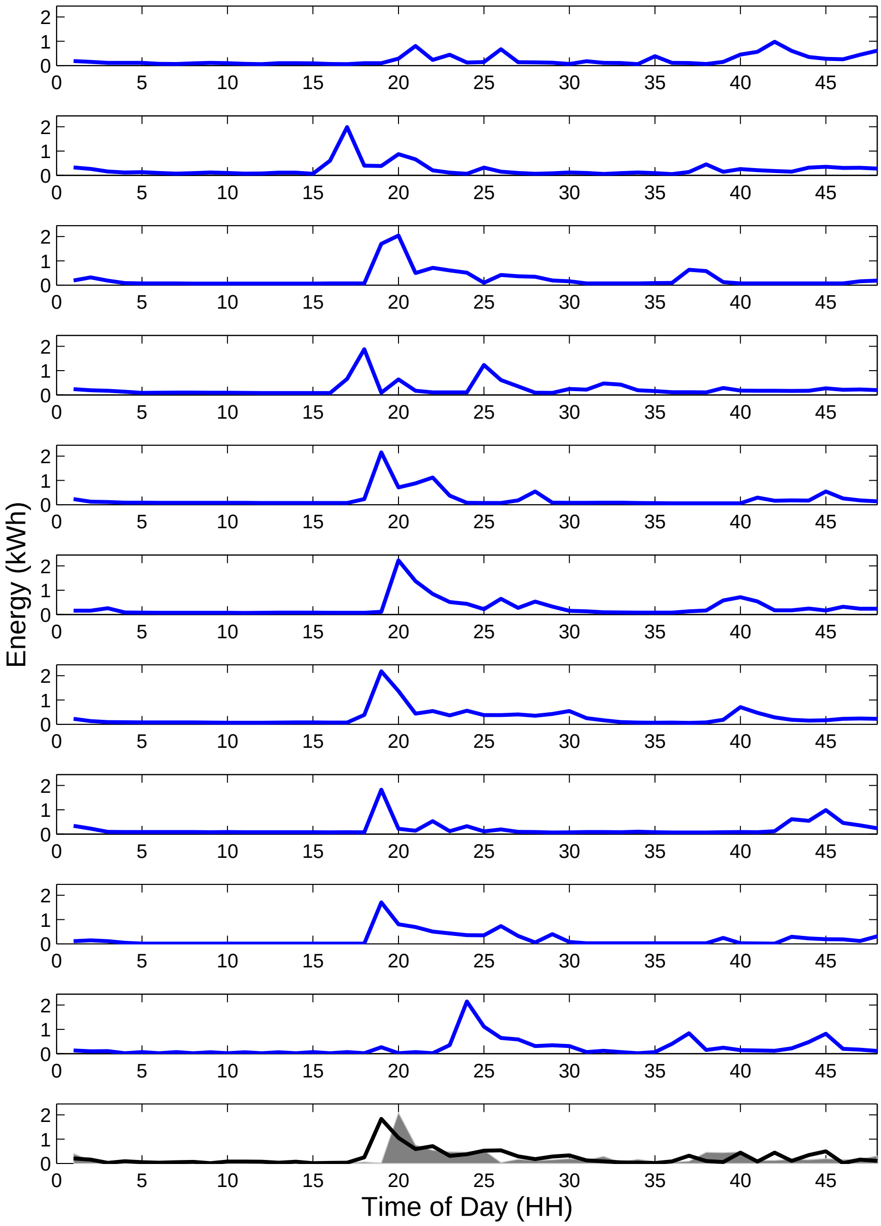

Figure 3: Forecast 236 (bold) against actual (shaded) with historical profiles (top 10 plots).

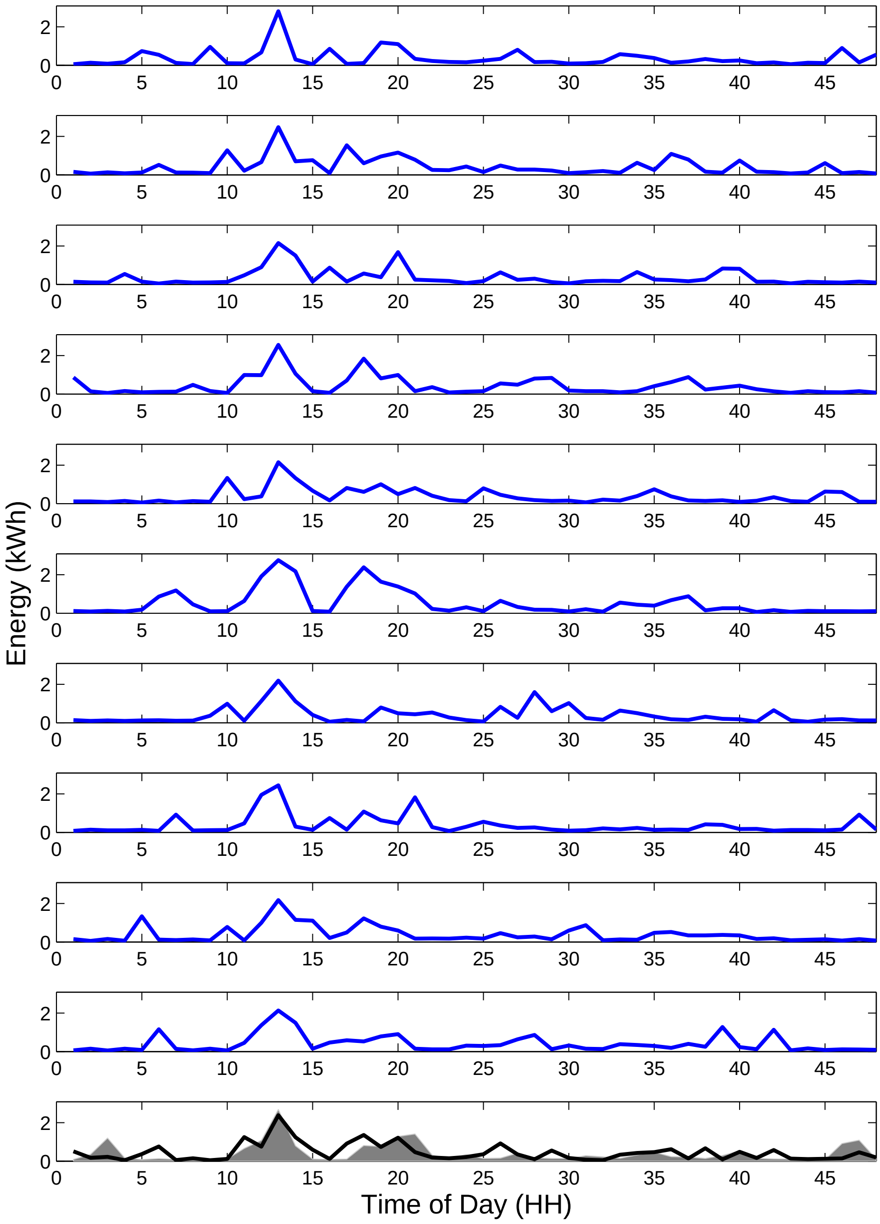

Figure 4: Forecast 268 (bold) against actual (shaded) with historical profiles (top 10 plots).

4 Discussion

In this paper we have developed an adaptable approach for forecasting spikey demand profiles by modelling a form of adjusted error measure first presented in Haben et al. (2014). In addition we have adapted the approach to bias early peaks which are desirable in many applications such as control of storage devices on the LV networks. We derived a maximum likelihood estimate (MLE) for our forecast which assumed an exponential model for the forecast errors with an age discounting factor on older data. We also derived a methodology for estimating the forecast and parameters for the MLE for our model and presented one possible way to model the prior information if no historical data was available. Finally we demonstrated our model for the forecasting of real smart meter data.

This paper serves as a first step to producing realistic forecasts of spikey data within a Bayesian framework. In future work refinements will be made to improve the model, including investigating the integration of weather variables which strongly affect domestic energy consumption.

Acknowledgements. The authors acknowledge financial support from the Thames Valley Vision project (SSET203 New Thames Valley Vision) via the Low Carbon Network Fund.

References

Charlton, N., Greetham, D. V., & Singleton, C. (2013). Graph-based algorithms for comparison and prediction of household-level energy use profiles. *International Workshop on Intelligent Energy Systems (IWIES) 2013*, .

Conn, A. R., Scheinberg, K., & Vicente, L. N. (2009). *Introduction to Derivative-Free Optimization (MPS-SIAM Series on Optimization)*. Society for Industrial and Applied Mathematics.

Department of Energy and Climate Change (DECC) (2013). Smart metering equipment technical specifications (v2.0). `https://www.gov.uk/government/consultations/smart-metering-equipment-technical-specifications-second-version`. [Online; accessed 31-Jan-2014].

Haben, S., Rowe, M., Greetham, D. V., Grindrod, P., Holderbaum, W., Potter, B., & Singleton, C. (2013). Mathematical solutions for electricity networks in a low carbon future. CIRED 22nd International Conference on Electricity Distribution, Stockholm, 10-13 June 2013.

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

Irish Social Science Data Archive (2012). Cer smart metering project, 2012. http://www.ucd.ie/issda/data/commissionforenergyregulation/. [Online; accessed 31-Jan-2014].

Keil, C., & Craig, G. C. (2009). A displacement and amplitude score employing an optical flow technique. *Weather and Forecasting*, *24*, 1297–1308.

Mitchell, M. (1998). *An Introduction to Genetic Algorithms (Complex Adaptive Systems)*. A Bradford Book; Third edition.

Molderink, A., Bakker, V., Bosman, M. G. C., Hurink, J. L., & Smit, G. J. M. (2010). A three-step methodology to improve domestic energy efficiency. IEEE Innovative Smart Grid Technologies Conference 19-21 Jan 2010.

Munkres, J. (1957). Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics*, *5*, 32–38.

Rowe, M., Holderbaum, W., Potter, B., & Liu, Y. (2012). The scheduling and control of storage devices on the low voltage network using forecasted energy demand. https://www.reading.ac.uk/CMOHB/resources/cmohbresources.aspx. [Online; accessed 30-Jan-2014].