Smart Meter Uncertainty Forecast

Estêvão Alvarenga[1,*], Jooyoung Jeon[1,*], Ran Li[2,*]

Claverton Down, Bath BA2 7AY, United Kingdom

Abstract

We propose to produce probabilistic forecasts for groups of households that minimise the forecast error, rather than individual household forecasts. This approach is also to be implemented in peer to peer energy trading strategies that aim to utilise surplus in renewable energy generation. We used two established methods for producing the forecasts: KDE and ARMA-GARCH. Our initial study indicates that probabilistic forecast accuracy is maximised with the aggregation size of 50 random customers or above. The research further suggests optimisation methods that allow smaller aggregation groups to have such a forecast accuracy, for different values of aggregated demand. For the optimal grouping we used a genetic algorithm approach, and compared the following different optimisation drivers: minimising standard deviation of the in-sample time-series, and minimising the forecast error of a cross-validation routine performed on both forecasting methods.

Keywords: Science, Publication, Complicated

1. Introduction

In the past, when conventional large power plants were responsible for most electricity supply, peak demand periods were estimated for large regions, aggregating many customers' data. The behaviour of the energy demand was understood by the distribution network and well modelled. Recently, with the aid of smart-grids and smart-metering devices, information with higher resolution is available per household. This new dimension of

*efsa@bath.edu

[1]University of Bath School of Management

[2]University of Bath Department of Electronic & Electrical Engineering

information could be used for enhancing the forecasts that support load-balancing decision-making by distribution companies, and to assess the feasibility of small scale investments, like electrical vehicles and domestic storage, in presence of a local energy marketplace. Recent approaches have forecasted the demand of each household, but it suffers high degree of uncertainty and high computation costs. In view of this, we propose to produce probabilistic forecasts for groups of households that minimise the forecast error, rather than individual households. The research further suggests optimisation methods that allow smaller aggregation groups to have such a forecast accuracy, for different values of aggregated demand.

Due to an increasing intermittent generation connections to the distribution network, accurate short-term forecasts for the electricity demand are crucial. System operators need accurate forecasts to balance their networks with other dispatchable generation and plan accordingly. Short-term demand forecasts are possible with smart meters and required for the efficient operation of the power distribution network, as well as for energy trading. The uncertain nature of individual household demand motivates the use of probabilistic forecasting. According to [1], the research interest in probabilistic energy forecasting has taken off rapidly in recent years.

Since we are producing short-term forecast (up to 24 hours), we focus on a time series model, rather than a regression models using weather data. History data is used to define a forecasting function, that relies on demand volume and load variation. One advantage of using time series data is the scalability and flexibility for implementing the same model in other regions or using other type of customers. It also does not depend on weather or socio-economical data availability.

Shorter forecast horizons, from seconds to several minutes, are used especially for load security (enhancement of frequency response and correction for generators forecasting accuracy) [2]. Forecasts from next-hour up to 7 days ahead can be used to capture value via intra-day or intra-week power arbitrage (moving energy from low value periods to high value periods). This horizon is also useful for planning optimal economic generation.

Generally, medium- and long-term forecasts must take into account the historical load and weather data, the number of customers in different categories, the appliances in the area and their characteristics including age, the economic and demographic data and their forecasts, the appliance sales data, and other factors [3]. This longer forecasts are used to predict loads as distant as twenty years ahead, so that expansion planning can be facili-

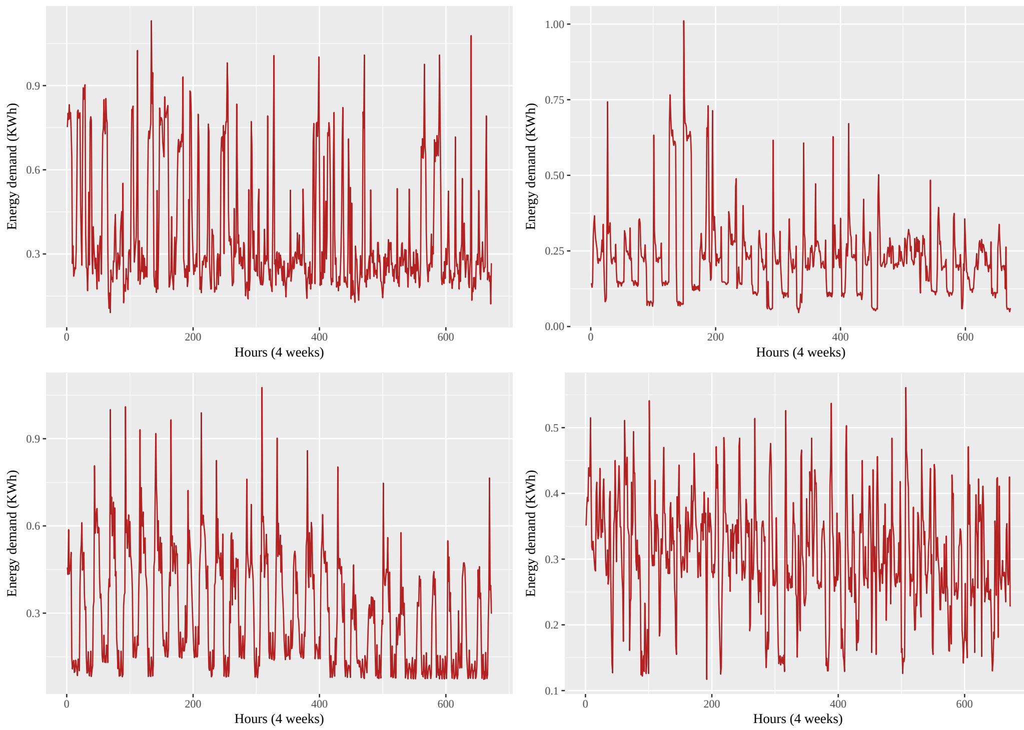

Figure 1: Example of demand pattern for different customers

tated. Large generation and infrastructure projects depend on this kind of forecasting for supporting decision [4].

2. The smart meter data

This section introduces the data used in the article, which corresponds to 1,000 residential smart meter data from Korea Electric Power Corporation (KEPCO). The time series contains the average energy used for each hour from January 2012 to October 2014. Figure 1 shows example of different behaviours during the same four week period for four different customers in February 2012. The values are expressed in KWh.

To avoid some periods of missing observations, we did not use the first 12 weeks of data. If there still was periods with missing observations, we used the following approach:

i. calculated the average from adjacent data for single missing observations;

ii. copied value from previous week, from same week day and hour, for multiple adjacent missing observations.

Periods of 12 weeks were used for model fitting and forecast generation for the next 24 hour. We implemented this technique for 168 different periods within the data set, with each of these periods starting 113 hours after the previous. We selected a 113-hour gap 168 times because this allows for the forecasts to start at all different hours at least once per each week in the dataset. This approach was necessary to enable a broad exploration while keeping computational time feasible.

2.1. Seasonality analysis

Before the application of the forecast methods, a seasonality analysis was performed as energy demand has natural cycles. The daily and weekly cycles are denoted as s_{01} and s_{02}. Since the smart meter data is recorded hourly, $s_{01} = 24$ and $s_{02} = 168$. The trend and s_{01} and s_{02} periodic patterns were removed from the data through the decomposition of the original signal in seasonal, trend and remainder components. This was performed using a seasonal-trend decomposition procedure based on Loess [5]. The annual seasonality, also naturally present in the original data, was left as part of the trend component. We chose this approach because all forecasting methods are using a few months of data, instead of years. The uncertainty forecasting methods were applied to the remainder component. The deterministic seasonality and trend were added to the density forecast before performance evaluations.

3. Forecasting methods

In this section, we briefly describe the two forecasting methods implemented: kernel density estimation (KDE) in Section 3.1 and autoregressive moving average generalised autoregressive conditional heteroskedasticity (ARMA-GARCH) in Section 3.2. We finish this section explaining our choice for the uncertainty forecast evaluation, the continuous ranked probability score (CRPS), in Section 3.3.

3.1. Unconditional kernel density estimation

Kernel density estimation (KDE) is a non-parametric method. This means it can maintain the original properties of the original distribution, avoiding previous assumptions of distributions. Similar to the unconditional KDE implementation done by [6], [7] and [8], the method enables the non-parametric estimation of a probability density f based on observations $\{Y_1, Y_2, ..., Y_n\}$. The unconditional KDE can be defined as Equation 1:

$$\hat{f}(y) = \sum_{w=1}^{n} K_{h_y}(Y_t - y) \tag{1}$$

where y is the energy demand forecast to be estimated, n is the size of the sliding window w, i.e. number of observations, and K is a Gaussian kernel function with bandwidth h_y.

The bandwidth is responsible for the density smoothness and was chosen according to Silverman's reference bandwidth, also known as Silverman's rule of thumb [9]. This method defines the bandwidth as Equation 2:

$$h = \begin{cases} 0.9\hat{\sigma}n^{-\frac{1}{5}} \ , & \text{if } \hat{\sigma} < \frac{sIQR}{1.34} \ , \\ 0.9\frac{sIQR}{1.34}n^{-\frac{1}{5}} \ , & \text{otherwise.} \end{cases} \tag{2}$$

where $\hat{\sigma}$ is the standard deviation of the sample and $sIQR$ is the sample interquartile range.

Similar to what was done by [7] different sliding window lengths were implemented. We observed better performance for shorter-term forecasting when smaller sliding windows were used (i.e. $w = 4$). Figure 2 examples how different customers have different energy demand density functions. Notice that the KDE was implemented on the remainder from the seasonality decomposition, that is why there are negative values.

3.2. Univariate ARMA-GARCH

According to [7], ARMA-GARCH models are used widely for capturing autocorrelation in the conditional mean and variance. The ARMA(p, q) describes the conditional mean process, while the GARCH(r, s) describes the conditional variance process. The GARCH(r, s) process is similar to an ARMA process implemented for the variance magnitude. The ARMA(p, q)-GARCH(r, s) model follows Equations 3

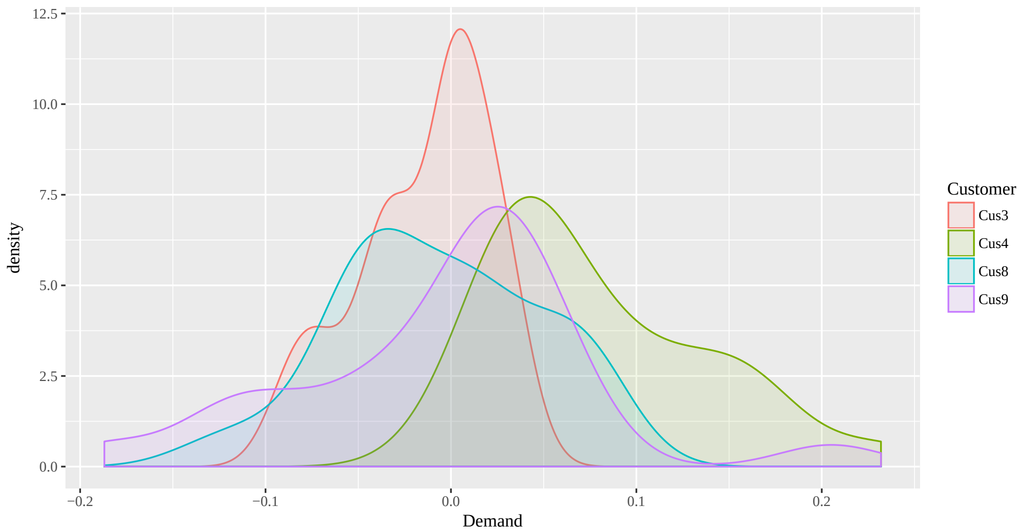

Figure 2: Example of different energy demand densities

$$y_t = \alpha_0 + \sum_{j=1}^{p} \alpha_j Y_{t-j} + \sum_{k=1}^{q} \beta_k \varepsilon_{t-k} \tag{3a}$$

$$\sigma_t^2 = \delta_0 + \sum_{l=1}^{r} \delta_j \sigma_{t-l}^2 + \sum_{m=1}^{s} \gamma_k \varepsilon_{t-m}^2 \tag{3b}$$

$$\varepsilon_t = \sigma_t \eta_t \tag{3c}$$

where y_t is the energy demand observed at time t; ϵ_t is a iid error term; σ_t is the conditional standard deviation (volatility); α_i, β_i, δ_i and γ_i are the coefficients of the AR, MA, GARCH and ARCH components with orders defined by non-negative integers p, q, r and s, respectively; and η_t is the white noise generating process. In principle the ε_t could follow any suitable probability distribution. For our model we chose the skewed version of generalised error distribution to allow flexibility for asymmetric data, avoiding normalisation.

The ARMA(p, q) order was defined via the lowest Bayesian Information Criterion (BIC) value for the combination of possible (p, q). BIC was described by [10] as a tool for choosing the appropriate dimensionality of a model, as an alternative to maximum likelihood and to Akaike Information Criteria (AIC) that leads to higher order models. The GARCH(r, s) order was set to (1,1).

3.3. Evaluation criteria

We used the continuous ranked probability score (CRPS) for the evaluation of uncertainty forecasts. The method, described in [11], assesses probabilistic forecasts of continuous variables that take the form of predictive densities or cumulative distributions. The CRPS can be considered a generalisation of the mean absolute error (MAE) to the case of probabilistic forecasts assessed against deterministic observations. The formal definition of this scoring method is described in Equation 4.

$$\begin{aligned} &\mathrm{CRPS} = \int_{-\infty}^{\infty} \mathrm{BS}(y) dy \\ &\text{where} \\ &\mathrm{BS}(y) = \frac{1}{T} \sum_{t=1}^{T} \{F_t(y) - \mathbf{1}(x_t \leq y)\}^2 \end{aligned} \tag{4}$$

The goal is the maximisation of the sharpness of the predictive distributions subject to calibration. While calibration refers to the statistical consistency between the predictive distributions and observations, sharpness is a forecast property that refers to the concentration of the predictive distributions.

3.4. Binary nonlinear optimisation

an using an integer optimisation, with domains [0,1] It runs 300 generations unless after 20 consecutive generations there is no enhancement on the function value. I implemented a 20 step optimisation, equally dividing the total demand in ranges. The optimal group was the one that, within the selected range, returned the lowest *decomposed noise deviation*.

Explain: genetic algorithm and stepwise frontier (bins) random? sdev? cvkd? cvag? **Heuristic?**

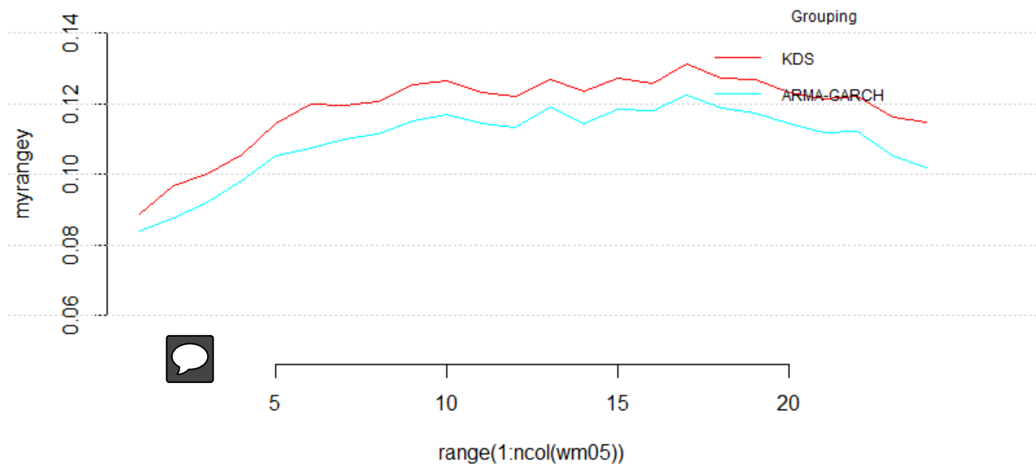

Figure 3: Benchmarking KDE Vs. ARMA-GARCH - **this picture will be updated**

4. Empirical results

As explained in Section 2, we produced individual household's post-sample forecasts from $t_1 = 1$ to $t_n = 60$ hours ahead, after each in-sample period of 12 weeks.

4.1. Benchmarking forecast methods

We averaged the CRPS values for 1,000 customers at 30 different sections from the data, from September 2012 to July 2013. We implemented two KDE models with different sliding windows sizes ($w_1 = 4$ and $w_2 = 24$), and one ARMA-GARCH model ($p, q \leq 3$ and $r, s = 1$). Figure 3 compares the forecast accuracy of selected methods.

4.2. Random grouping forecasts

We grouped the households randomly in different group sizes, and produced new forecasts for each of the groups. Each forecast was evaluated against post-sample data and averaged on the group size domain. Figure 4 summarises the result for groups containing up to 50 customers. The

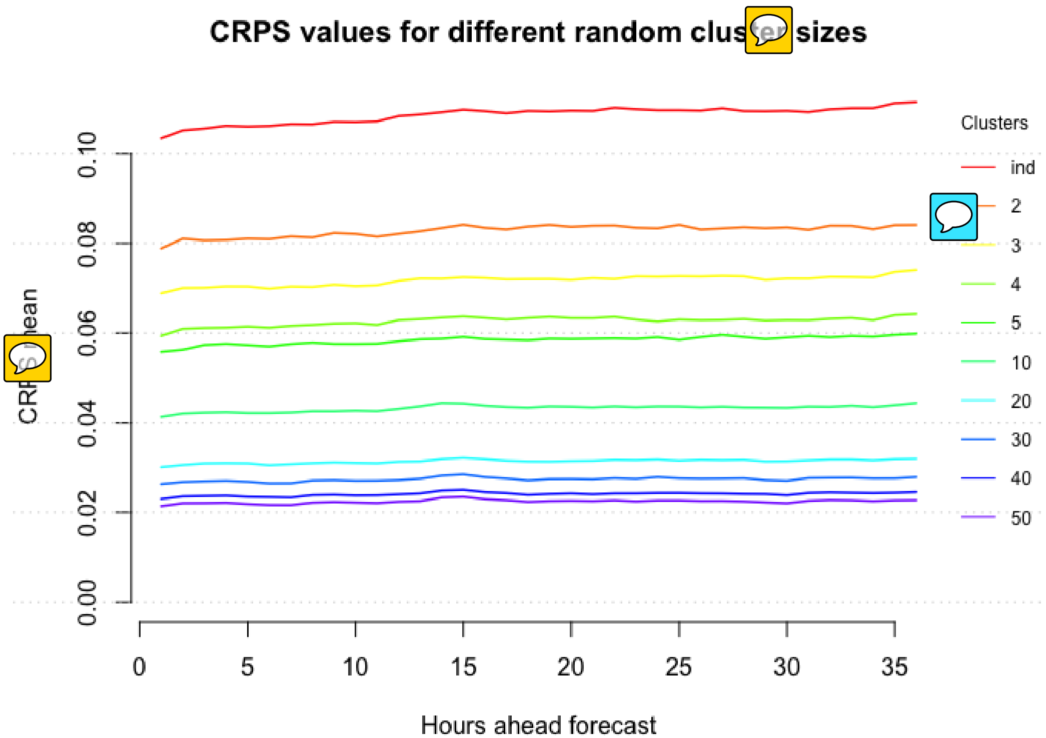

Figure 4: Forecasts for random groups - **this picture will be updated**

accuracy of the forecast is expressed by it's CRPS value (the lower the better). There is a natural tendency of worse accuracy as the forecast horizon increases. The converging result suggests that larger groups would only marginally enhance the forecast accuracy.

4.3. Optimal grouping forecasts

After realising the enhancements are marginal while increasing the random group sizes, we implemented a few grouping techniques: one based on standard deviation minimisation, and one based on cross-validation on the forecast error. We used the standard deviation as the statistical dispersion measure. As described in Section 3.4 the genetic algorithm searched, for each bin in the stepwise frontier, the best combination of households that would minimise the standard deviation of the remainder, using deseasonalised aggregated demand data.

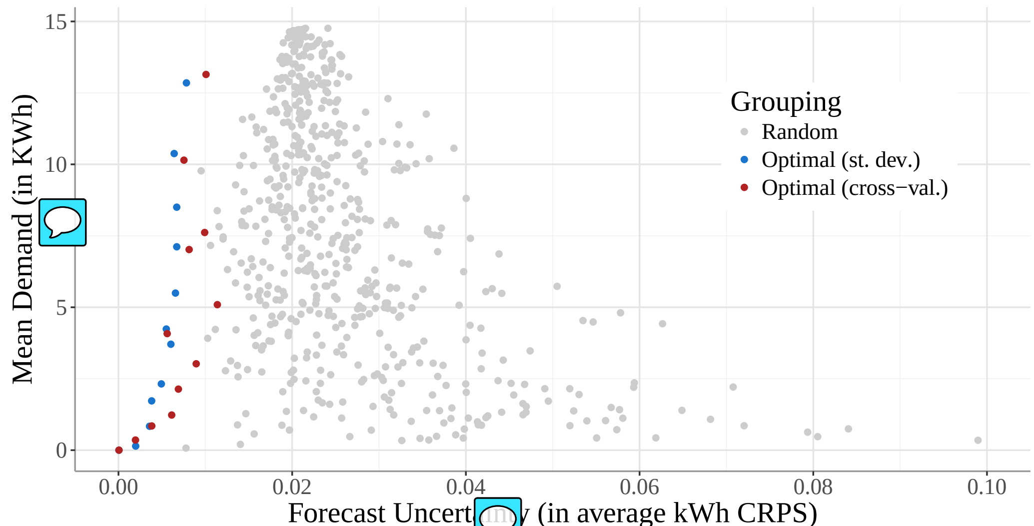

Figure 5: Optimal grouping of customers

For the cross-validation grouping, we divided the last week from in-sample in 32 sections. An ARMA-GARCH forecast was produced for each of these sections and evaluated against the real demand data via CRPS. The group that returned the minimal CRPS average for the chosen forecast horizon was selected for each bin in the stepwise frontier.

For a given bin, the standard deviation grouping delivers the same group irregardless of the forecast horizon. On the other hand, the cross-validation depends on what forecast horizon error is to be minimised. In respect to that, a group created for a $t_{f1} = 12$ forecast won't be the best grouping for a $t_{f2} = 4$. Figure 5 exemplifies this procedure.

References

[1] 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.

[2] H. Al-Hamadi, S. Soliman, Short-term electric load forecasting based on kalman filtering algorithm with moving window weather and load model, Electric Power Systems Research 68 (2004) 47 – 59.

[3] E. A. Feinberg, D. Genethliou, Load forecasting, in: Applied mathematics for restructured electric power systems, Springer, 2005, pp. 269–285.

[4] S. Soliman, S. Persaud, K. El-Nagar, M. El-Hawary, Application of least absolute value parameter estimation based on linear programming to short-term load forecasting, International Journal of Electrical Power & Energy Systems 19 (1997) 209–216.

[5] R. B. Cleveland, W. S. Cleveland, I. Terpenning, Stl: A seasonal-trend decomposition procedure based on loess, Journal of Official Statistics 6 (1990) 3.

[6] S. Arora, J. W. Taylor, Forecasting electricity smart meter data using conditional kernel density estimation, Omega 59, Part A (2016) 47 – 59. Business Analytics.

[7] J. Jeon, J. W. Taylor, Short-term density forecasting of wave energy using arma-garch models and kernel density estimation, International Journal of Forecasting 32 (2016) 991 – 1004.

[8] J. W. Taylor, J. Jeon, Forecasting wind power quantiles using conditional kernel estimation, Renewable Energy 80 (2015) 370 – 379.

[9] B. W. Silverman, Density estimation for statistics and data analysis, volume 26, CRC press, 1986.

[10] G. Schwarz, Estimating the dimension of a model, Ann. Statist. 6 (1978) 461–464.

[11] T. Gneiting, F. Balabdaoui, A. E. Raftery, Probabilistic forecasts, calibration and sharpness, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69 (2007) 243–268.