Forecast-based portfolio optimisation of electricity demand for P2P markets

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 enhances with customer aggregation, converging to an average maximum for groups larger than 50 customers. The research further suggests optimisation methods that allow smaller aggregation groups to have such a forecast accuracy, for different values of aggregated demand. We implemented a genetic algorithm optimisation, based on portfolio selection theory, for defining groups of customers with minimal expected forecast errors. The expected forecast error function used three different approaches, and we found the simplest of them performing as good as the others. enhance abstract after paper is better defined

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. Many customers' data was interpreted in an aggregated fashion, and

*efsa@bath.edu

[1]University of Bath School of Management

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

the overall energy demand behaviour was understood by the distribution network. Electric utilities have always forecast the hourly aggregated loads as well as peak loads to schedule generator maintenance and to choose an optimal mix of on-line capacity [1]. Recently, the need for accurate high frequency forecasts is even greater as the landscape of the energy industry is changing. The reason is the fast pace adoption of breakthrough technologies like distributed generation, electric vehicles (EV), and energy storage [2].

A proportion of generation will come from more local sources, including medium to small community biomass-fuelled power plants, locally generated waste, and local wind sources. These will feed local distribution networks that can sell excess capacity into the grid [3]. In addition, other renewable energy sources based on solar, geothermal and tides, could be used to meet a large portion of the energy demand. However, most of these resources are not actively utilised at distribution systems, as there are no incentive from an active local energy marketplace. A distributed approach to system design is proposed by [2] based on a peer-to-peer (P2P) electricity trading and control model.

The authors indicate that a P2P philosophy is well suited and even preferred when a large number of small scale distributed energy resources are connected to the grid. The envisioned P2P energy platform suggests the P2P trading will take place in the operation of the spare capacity of the distribution network. Actors responsible for procuring aggregation of demand, combining multiple short-duration consumer loads, and trading on organised energy market auctions, will require suitable demand and generation forecasts.

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 [4]. This longer forecasts are used to predict loads as distant as twenty years ahead, so that expansion planning can be facilitated. Large generation and infrastructure projects depend on this kind of forecasting for supporting decision [5]. Short-term electric load forecasting, on the other hand, is vital for power generation and operation. It is fundamental in many applications such as providing optimal economic generation, system security, and management and planning [6]. It can also be used to capture value via intra-day or intra-week power arbitrage (moving energy from low value periods to high value periods).

While the decision making process in the utility industry rely on expected values, or point forecasting, the increase in market competition and renewable integration requires a probabilistic approach for planning and operation of energy systems [7]. The research interest in probabilistic energy forecasting has taken off rapidly in recent years [8] and, at the same time, a massive smart meter deployment providing the industry a huge amount of high resolution data [7].

Our research proposes a methodology to support trade within the envisioned P2P energy marketplace through demand aggregation based on minimised forecast error. History data, coupled with recent development on probabilistic forecast and computer power, is used to define a time series based forecasting function, that relies on demand volume and load variation. One advantage of using time series data is the simplicity 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, as it uses past trend and behaviour for extrapolating the future energy requirement [9]. Wherever there is additional datasets available, regression based models can be used. This have been proved to enhance the prediction accuracy in noisy financial time series forecasting [10]. enhance intro after paper is better defined

The rest of the paper is organised as follows: ...

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). We had no access to specific location data for security purposes. 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.

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.

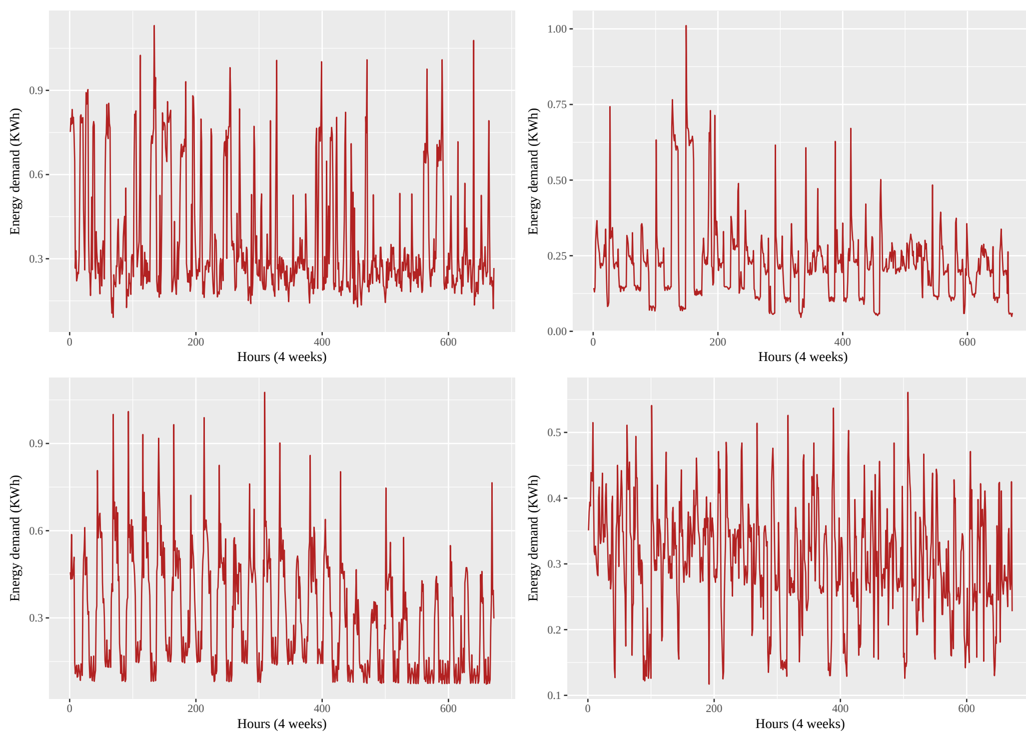

Figure 1: Example of demand pattern for different customers

Datasets containing 12 weeks were used for model fitting and forecast generation for the next 24 hour. From the full data we defined 168 different datasets. Each of these datasets were 113 hours apart, as this allowed for the forecasts to start at all different hours inside weeks, at least once.

2.1. Seasonal analysis

Some time-series methods are capable to deal with seasonality, however [11] evidenced that deseasonalising prior to modelling improves the forecast accuracy of traditional statistical methods. check in book library Location-Library Level 5 658.012.2 WHE For this reason, before the application of the forecast methods, a seasonal analysis was performed as energy demand has natural cycles.

The oldest approach to handling seasonality in time series is to extract it using a seasonal decomposition procedure such as the X-11 method. In addition to work on the X-11 method and its variants, there have also been several new methods for seasonal adjustment developed, the nonparametric method STL being one of them [12].

STL (seasonal-trend decomposition based on Loess) addresses the main drawbacks from X-11, without letting go of its innovations. X-11 has a complex option selection procedure, is difficult to diagnose, and is not prepared to deal with missing data. The more recent STL method does not suffer from these, while delivering a robust seasonal and trend-cycle decomposition that is not distorted by transient, aberrant behaviour in the data [13].

[2do] [efsa] do I need to explain the STL mathematically?

Because of its flexibility, we implemented the STL seasonal decomposition, before running any forecasting model. The daily and weekly patterns 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-cycle and s_{01} and s_{02} periodic patterns were estimated and a decomposed signal was generated.

The annual seasonality, also present in the original data, was estimated as part of the trend-cycle signal. We chose this approach because all forecasting methods are using a few months of data, and an yearly seasonal pattern can be interpreted as a longer cycle within the data. The uncertainty forecasting methods were applied to the residual component, after removing seasonal and trend-cycle components. In 3.3 we will describe the evaluation criteria for the density forecasts we produced, and how we treated the removed componentes.

3. Forecasting methods

In this section, we 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. In 3.3, we will describe the evaluation criteria for the density forecasts we produced.

3.1. Unconditional kernel density estimation

Kernel density estimation (KDE) is a non-parametric method. This means it can maintain the original properties, avoiding any previous assumptions of distributions, constructing the density function based on historic observations. Similar to the unconditional KDE implementation done by [14], [15] and [16], 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 through 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 [17]. This method defines the bandwidth through 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 sliding window w and $sIQR$ is the sample interquartile range.

Similar to what was done by [15], different sliding window lengths were implemented.

[2do] [efsa] should I describe different sliding windows (4,6,8,12,24) with a picture?

We observed better performance for shorter-term forecasting when smaller sliding windows were used (i.e. $w = 4$). Figure 2 examples how different the

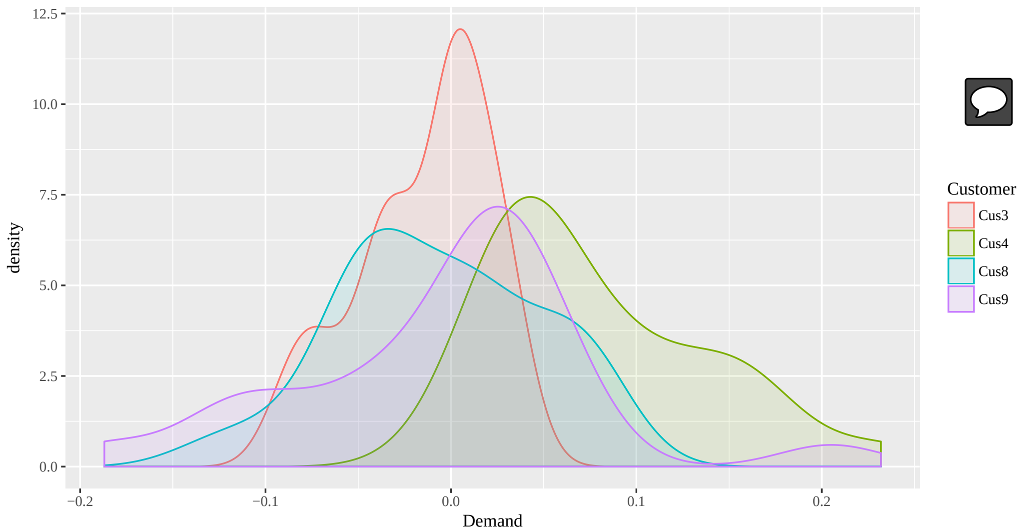

Figure 2: Example of different energy demand densities

unconditional distribution functions of different customers could be. As described in 2.1 the KDE was implemented on the residuals from the seasonal decomposition. This is why there are negative values.

3.2. Univariate ARMA-GARCH

According to [15], ARMA-GARCH models are used widely for capturing autocorrelation in the conditional mean and variance. GARCH models were used to predict day-ahead electricity prices in Spain and California, outperforming general time series ARIMA models, specially when high volatility is present [18]. 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.

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

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

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

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 [19] 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).

[2do] [efsa] should we explain we used up to ARMA(3,3) with best BIC? Should we explain how we got to (3,3) instead of (10,10) for example?

[2do] [efsa] should we explain the bad ARMA-GARCH fit using KDE? And explain how we got to the best gof.min? we have a plot for this!

3.3. Evaluation criteria

With the forecasts in hands, we accordingly added back the deterministic seasonality removed previously, as described in **??**. The trend-cycle component was extrapolated with the last value recorded in the in-sample data. We then used the continuous ranked probability score (CRPS) for the evaluation of uncertainty forecasts. The method, described in [20], 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. Benchmarking forecast methods

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

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, as defined in Section 3.1, with different sliding windows sizes ($w_1 = 4$ and $w_2 = 24$); and one ARMA-GARCH model ($p, q \leq 5$ and $r, s = 1$), as defined in Section 3.2. Figure 3 compares the forecast accuracy of the two methods.

4. Demand Aggregation

One of the foreseen roles in the envisioned P2P energy marketplace is the P2P aggregator. This actor will deliver, among other services, demand and generation forecast for aggregated load [2]. For a given aggregated generation, a group of customers has to be selected whose forecasted demand closely match the generation. Within the distribution network, this group could comprise from one to all customers.

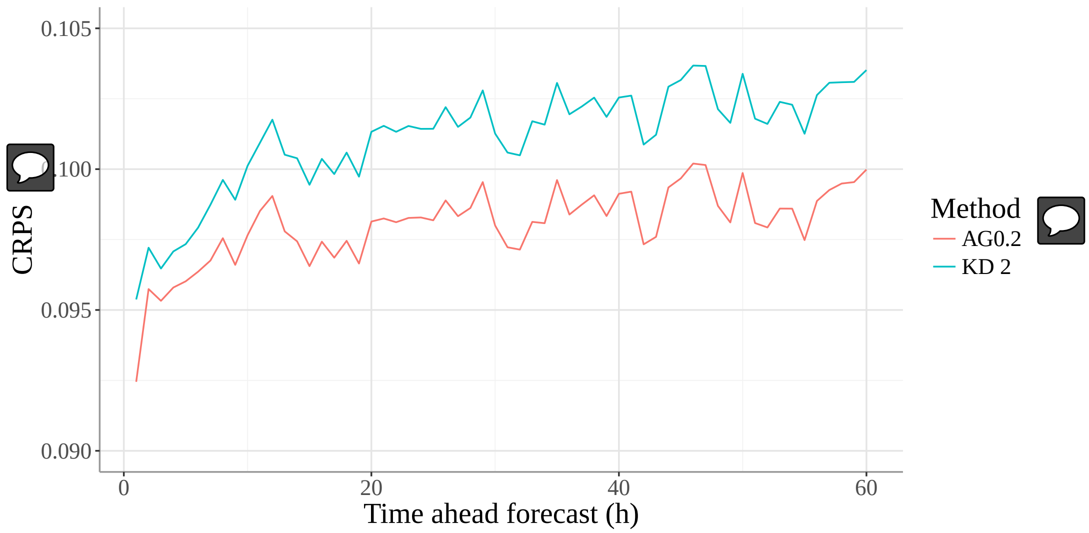

Figure 3: Benchmarking KDE Vs. ARMA-GARCH

4.1. Random grouping forecasts

In order to understand the relationship between demand and forecast uncertainty, we aggregated the demand of households chosen randomly in various group sizes, and produced new forecasts for each group. In order to maintain proportional comparability, the aggregated demand value of each group is the sum of its households' demand divided by the number of households. Each forecast was evaluated against post-sample data and averaged for similar sized groups. Figure 4 summarises the result for groups containing up to 50 customers. Further empirical tests indicate that 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.

4.2. Cardinality constrained portfolio optimisation

From the random grouping results, we can observe the aggregated demand generate forecasts with a lower errors. We hypothesised that a selection aiming at minimising the forecast error for a certain range of aggregated demand, would be able to further enhance the forecast ability. These optimal groups, if feasible, would be able to match comunal generation at different aggregation levels. To define the best customer grouping for different aggregated

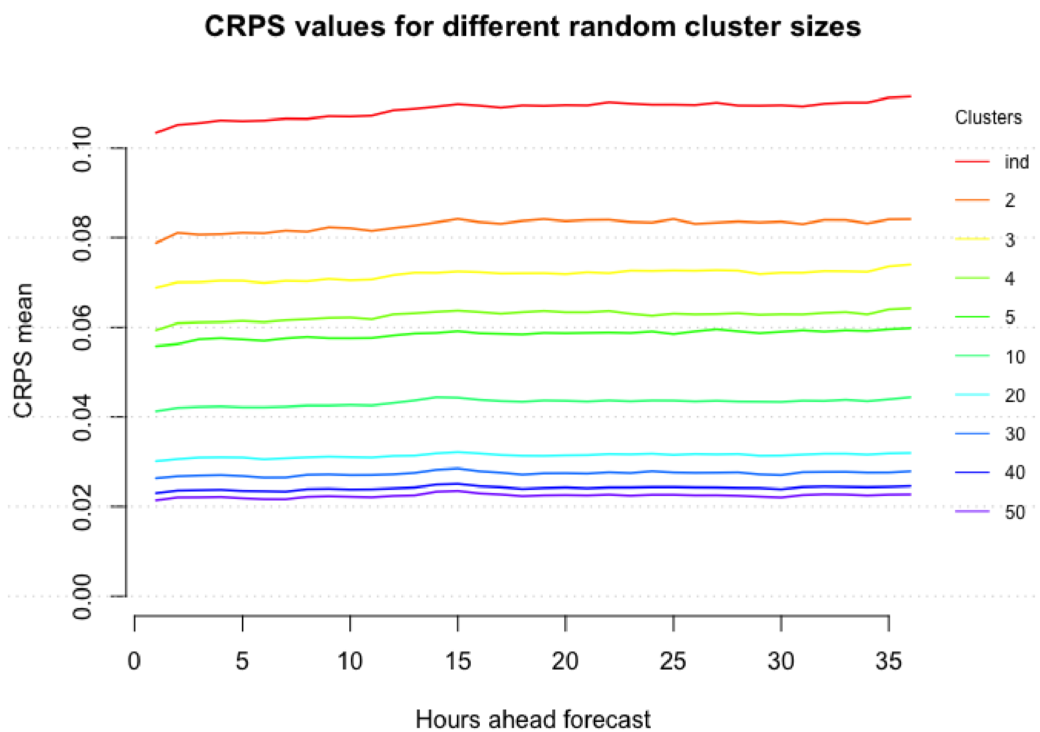

Figure 4: Forecasts for random groups - we have to think of alternative fig

demand values is similar to designing a risk-return plot from classic portfolio optimisation theory.

Markowitz's portfolio selection approach studies how investors can construct optimal portfolios, taking into consideration the trade-off between market volatility and expected returns. For a particular universe of assets, portfolios on the efficient frontier offer the maximum possible expected return for a given level of risk. One of the several formulations involves the construction of a portfolio with minimal risk provided that a prescribed return level is attained [21]. In the P2P energy marketplace scenario, risk and return can, respectively, be interpreted as the expected forecast error of each group of customers and expected aggregated demand. To construct the efficient frontier in this scenario we need to select groups that, for a given expected aggregated demand returns the lowest expected forecast error.

Classic portfolio optimisation assumes that asset returns follow a multivariate normal distribution. This means that the return on a portfolio of assets can be completely described by the expected return and the variance. The portfolios on the efficient frontier can be found by quadratic programming, a widely available and efficient solver. The solutions are optimal and the selection process can be constrained by practical considerations which can be written as linear constraints [22].

However, energy demand for different customers does not behave like assets for a few reasons:

i. The perishable nature of electricity demand;
ii. Cardinality restrictions, as customers either participate to a group or not and;
iii. Nonlinear relationship between customer selection and forecast error.

Retrieving the expected aggregated demand is trivial, considering the before mentioned forecasting techniques. On the other hand, minimising the expected forecast error can be performed in different ways. We chose three main approaches on our work, described in the next sections:

i. Minimising standard deviation of the residual of a deseasonalised aggregated demand (SD-residual);
ii. Minimising the deviation of the seasonal pattern of customers pertaining a group and (SD-seasonal);
iii. Minimising forecast error using the forecasting function on a validation dataset (Validation).

As explained, all three approaches require nonlinear functions f_i to be applied to a customer selection vector v. Considering the cardinality restrictions, v was defined as a binary vector with as many positions as the universe of available customers. We chose a genetic algorithms approach as this have been proven to perform marginally better than tabu search and simulated annealing for cardinality constrained portfolio optimisation problems. The attraction are the effectively independency of the selected objective function, and the feasibility of integer constraints is adoption [22].

should we describe the algorithm, mathematically?

4.3. Minimisation of SD on residual

I am 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*. We used the standard deviation as the statistical dispersion measure.

Explain: stepwise frontier (bins) As described the genetic algorithm searched, for each bin in the stepwise frontier, the best combination of households that would minimise the standard deviation of the residual, using deseasonalised aggregated demand data.

4.4. Minimisation of SD on seasonal signal

From $\boldsymbol{S}_{Cus\times168}$ I get the partition $\boldsymbol{S'}_{SelectCus\times24}$ where $SelectCus$ are only the rows of the customers assigned to that group. I then calculate the standard deviation of each column of $\boldsymbol{S'}$ and then the mean of all results. The difference between the models are the combination of $\boldsymbol{S}$ being scaled or not, before any operation, and the partition including 24 columns or just 1.

we have 2 plots comparing seas and sdev, and a combination of them, should we plot? - will look the same plot for comparing the cval

4.5. Minimisation of forecast error on validation dataset

For the validation grouping approach to finding the optimal aggregation, we divided the last week of the in-sample data into 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. Figure 5 exemplifies the results of this procedure.

there is a new comparison plot with more options, but less clear

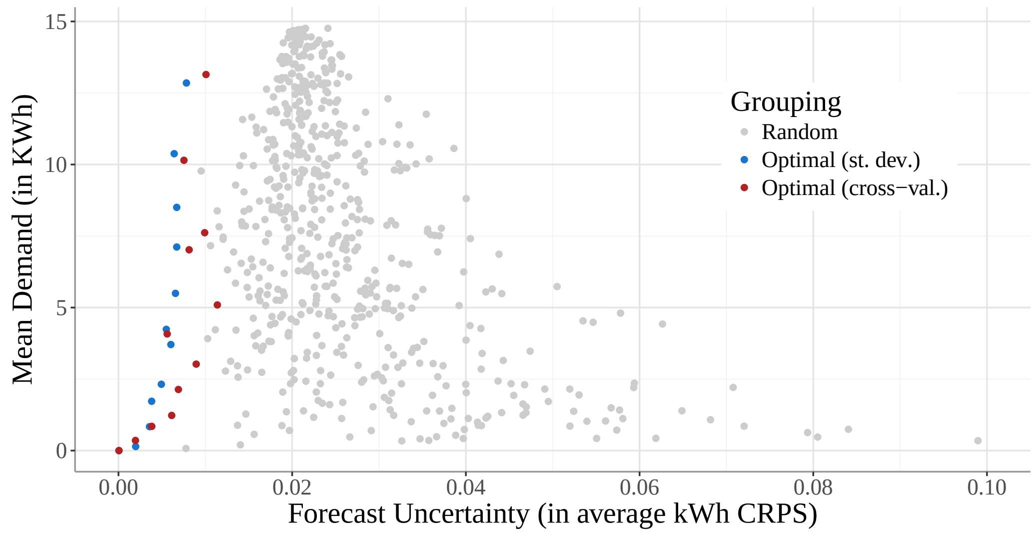

Figure 5: Optimal grouping of customers

4.6. Benchmark optimal aggregation

For a given bin, the standard deviation grouping assumes the same group regardless of the forecast horizon. On the other hand, the validation approach allows us to find the optimal composition of households on a certain forecast horizon of our interest. In respect to that, a group created for a $t_{f1} = 12$ forecast won't be the best grouping for a $t_{f2} = 4$. where should we talk about relaxation of sdev?

discuss findings: overfitting, etc...

References

[1] R. Ramanathan, R. Engle, C. W. Granger, F. Vahid-Araghi, C. Brace, Short-run forecasts of electricity loads and peaks, International Journal of Forecasting 13 (1997) 161 – 174.

[2] A. Pouttu, J. Haapola, P. Ahokangas, Y. Xu, M. Kopsakangas-Savolainen, E. Porras, J. Matamoros, C. Kalalas, J. Alonso-Zarate, F. D. Gallego, et al., P2p model for distributed energy trading, grid control and ict for local smart grids, in: Networks and Communications (EuCNC), 2017 European Conference on, IEEE, 2017, pp. 1–6.

[3] J. Hain, G. Ault, S. Galloway, A. Cruden, J. McDonald, Additional renewable energy growth through small-scale community orientated energy policies, Energy Policy 33 (2005) 1199 – 1212.

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

[5] 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.

[6] 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.

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

[8] 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.

[9] L. Suganthi, A. A. Samuel, Energy models for demand forecasting—a review, Renewable and Sustainable Energy Reviews 16 (2012) 1223 – 1240.

[10] F. E. Tay, L. Cao, Application of support vector machines in financial time series forecasting, Omega 29 (2001) 309 – 317.

[11] S. C. Wheelwright, S. G. Makridakis, et al., Forecasting methods for management, Wiley, 1973.

[12] J. G. D. Gooijer, R. J. Hyndman, 25 years of time series forecasting, International Journal of Forecasting 22 (2006) 443 – 473. Twenty five years of forecasting.

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

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

[15] 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.

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

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

[18] R. C. Garcia, J. Contreras, M. Van Akkeren, J. B. C. Garcia, A garch forecasting model to predict day-ahead electricity prices, IEEE transactions on power systems 20 (2005) 867–874.

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

[20] 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.

[21] P. Bonami, M. A. Lejeune, An exact solution approach for portfolio optimization problems under stochastic and integer constraints, Operations Research 57 (2009) 650–670.

[22] T.-J. Chang, N. Meade, J. Beasley, Y. Sharaiha, Heuristics for cardinality constrained portfolio optimisation, Computers & Operations Research 27 (2000) 1271 – 1302.