

Forecast-based portfolio optimisation of electricity demand for P2P markets

Estêvão Ferreira Sêco de Alvarenga^{1,*}, Jooyoung Jeon^{1,*}, Fotios Petropoulos^{1,*}, Ran Li^{2,*}, Furong Li^{2,*}

University of Bath, UK

Abstract

We propose to produce probabilistic forecasts for a group of households rather than individual households that minimise forecast uncertainty. This approach is also useful for peer to peer (P2P) energy trading that aims to utilise surplus in renewable energy generation locally. To produce the probabilistic forecasts, we used two probabilistic forecast models: Kernel density estimation (KDE) and ARMA-GARCH. Our study indicates that probabilistic forecast accuracy enhances with a larger aggregation of customers, and converges to an average maximum for groups larger than 50 customers (I am trying to move away from number of customers, and more towards demand aggregation), the number of minimum customers required in P2P when the aggregation is constructed in a random manner. The research further suggests methods of defining a group of customers, which allows to have a better probabilistic forecast accuracy, for an expected aggregated demand given. Among the three different methods we explored, the simplest method using

*efsa@bath.edu

¹University of Bath School of Management

²University of Bath Department of Electronic & Electrical Engineering

standard deviation in the in-sample period performs as good as the others. Our empirical analysis uses hourly data from Korea and Ireland to evaluate density forecasts, up to 24 h ahead. Check Irish data/result is included. Any further findings?

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 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 (Ramanathan et al., 1997). 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 (Pouttu et al., 2017).

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 (Hain et al., 2005). 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 Pouttu et al. (2017) 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 (Feinberg and Genethliou, 2005). These 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 (Soliman et al., 1997). 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 (Al-Hamadi and Soliman, 2004). 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 (Hong and Fan, 2016). The research interest in probabilistic energy forecasting has taken off rapidly in recent years (Hong et al., 2016) and, at the same time, a massive smart meter deployment providing the industry a huge amount of high resolution data (Hong and Fan, 2016).

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 (Suganthi and Samuel, 2012). 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 (Tay and Cao, 2001). 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

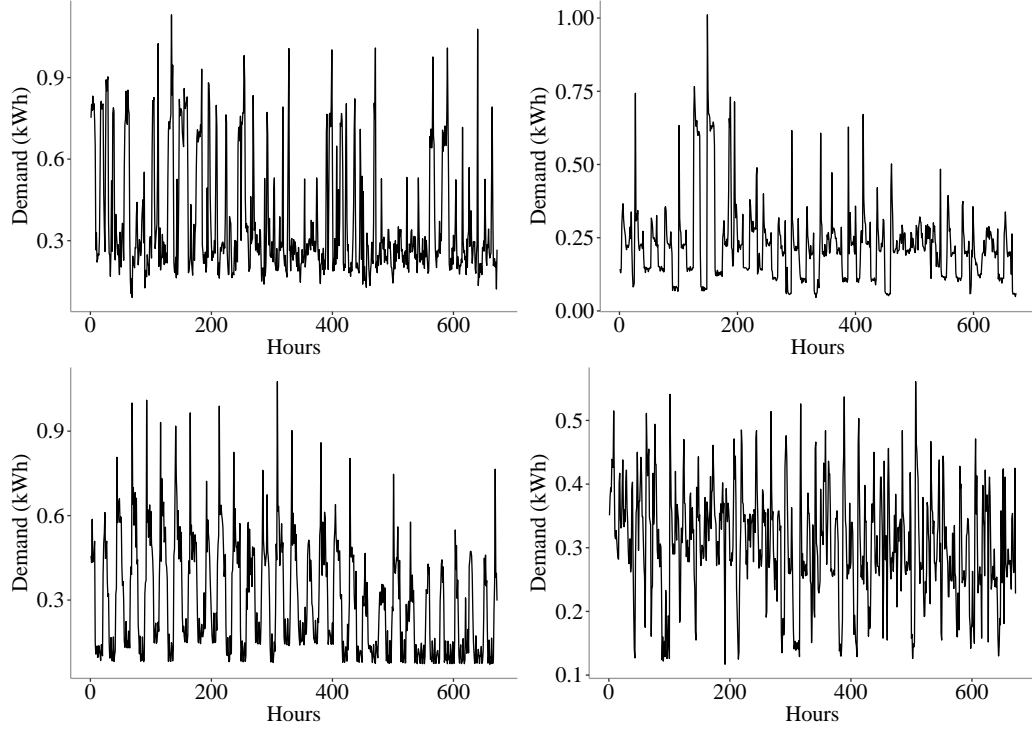


Figure 1: Example of electricity demand pattern for four different customers during a period of four weeks

(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.

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 Wheelwright et al. (1973) evidenced that deseasonalising prior to modelling improves the forecast accuracy of traditional statistical methods. [check in book library LocationLibrary 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 (Gooijer and Hyndman, 2006).

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 (Cleveland et al., 1990).

STL is a filtering procedure for decomposing the original data Y_t into trend-cycle T_t , seasonal S_t and residual R_t (1), using a local regression technique known as Loess.

$$Y_t = T_t + S_t + R_t \quad (1)$$

Loess delivers distance-weighted least-squares fit, of polynomials of degree d , to localised subsets of q elements from the data containing n elements. Large q results in traditional smoother regressions, while $q \rightarrow \infty$ generates ordinary polynomial fits. STL consists on two procedures, updating seasonal and trend components, while computing robustness weights recursively. This was implemented using the R `stats` package (R Core Team, 2016), 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 components.

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 Arora and Taylor (2016), Jeon and Taylor (2016) and Taylor and Jeon (2015), the method enables the non-parametric estimation of a probability density f based on observations $\{Y_1, Y_2, \dots, Y_n\}$. The unconditional KDE can be defined through Equation 2:

$$\hat{f}(y) = \sum_{w=1}^W K_{h_y}(Y_w - y) \quad (2)$$

where y is the energy demand forecast to be estimated, W 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 (Silverman, 1986). This method defines the bandwidth through Equation 3:

$$h = \begin{cases} 0.9\hat{\sigma}n^{-\frac{1}{5}} , & \text{if } \hat{\sigma} < \frac{sIQR}{1.34} , \\ 0.9\frac{sIQR}{1.34}n^{-\frac{1}{5}} , & \text{otherwise.} \end{cases} \quad (3)$$

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 Jeon and Taylor (2016), 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 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 Jeon and Taylor (2016), 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 (Garcia et al., 2005). 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 4.

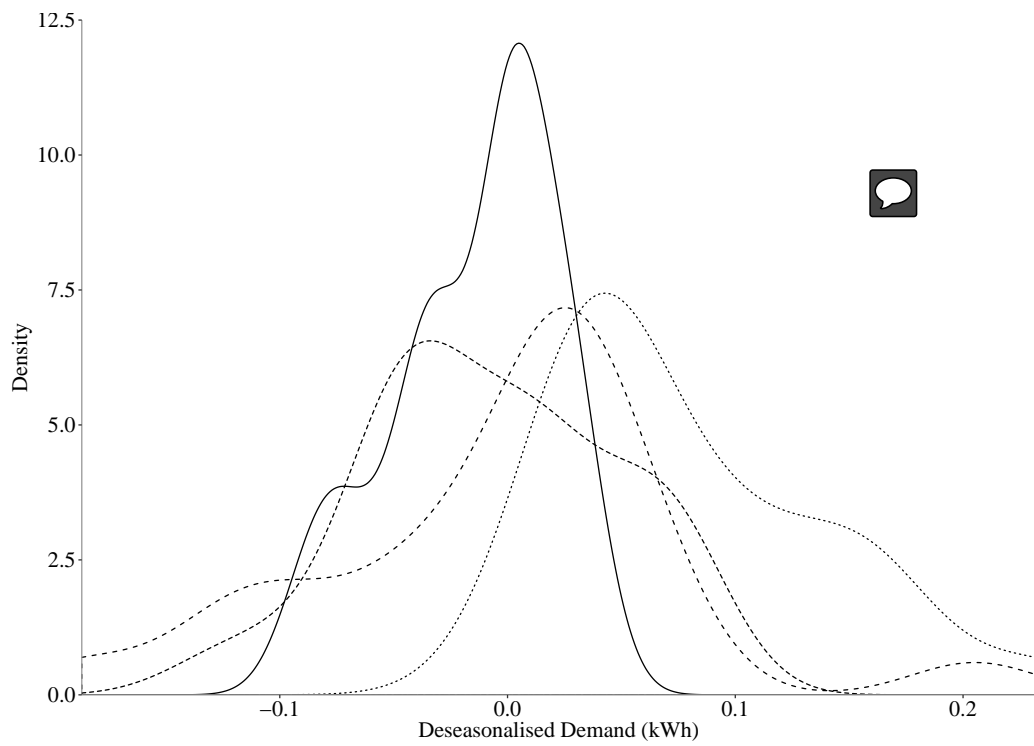


Figure 2: Examples of electricity demand densities forecasted for four different individual customers, built using KDE with a window size of 24h

$$y_t = \alpha_0 + \sum_{j=1}^p \alpha_j Y_{t-j} + \sum_{k=1}^q \beta_k \varepsilon_{t-k} \quad (4a)$$

$$\sigma_t^2 = \delta_0 + \sum_{l=1}^r \delta_l \sigma_{t-l}^2 + \sum_{m=1}^s \gamma_m \varepsilon_{t-m}^2 \quad (4b)$$

$$\varepsilon_t = \sigma_t \eta_t \quad (4c)$$

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 Schwarz (1978) 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 (G)? 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 2.1. 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 Gneiting et al. (2007), 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 5.

$$\text{CRPS} = \int_{-\infty}^{\infty} \text{BS}(y) dy$$

where

$$\text{BS}(y) = \frac{1}{T} \sum_{t=1}^T \{F_t(y) - \mathbf{1}(x_t \leq y)\}^2 \quad (5)$$

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

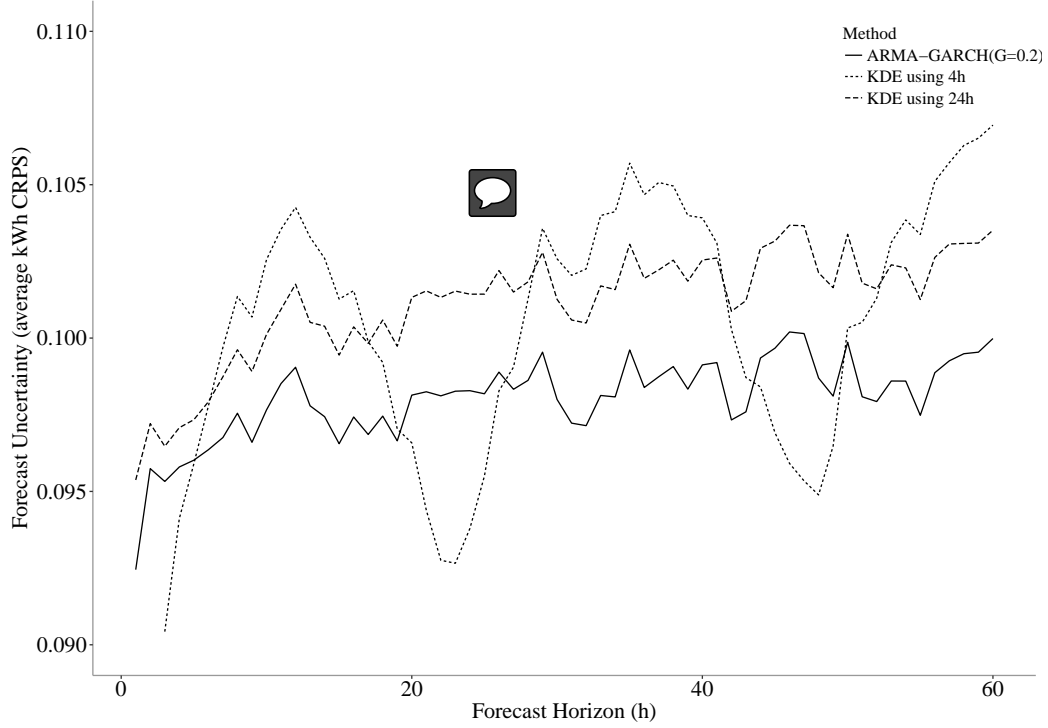


Figure 3: Benchmarking forecasting methods: KDE with window sizes of 4h and 24h, and ARMA-GARCH using $gofmin = 0.2$

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 (Pouttu et al., 2017). 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.

4.1. Random grouping forecasts

In order to understand the relationship between demand aggregation and forecast uncertainty, we created several groups containing a random selection of households, within a population of 100 households. Each of these groups had their demand forecasted using the best ARMA-GARCH model. The forecast was evaluated against post-sample data using CRPS, described on section 3.3. 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.

The results were summarised through similar expected demand groups (Figure 4). The converging pattern 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,

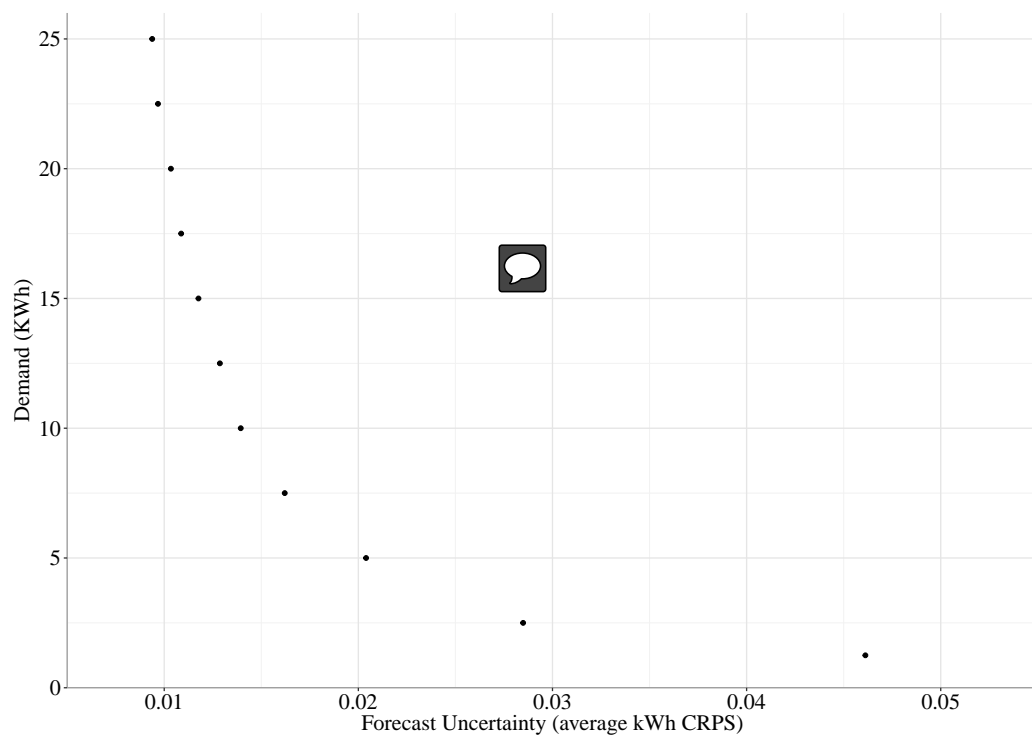


Figure 4: Forecast uncertainty for different sizes of random groups of households



if feasible, would be able to match communal generation at different aggregation levels. To define the best customer grouping for different aggregated 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 (Bonami and Lejeune, 2009). 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 (Chang et al., 2000).

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 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 households. We chose a genetic algorithm 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 selection (Chang et al., 2000).

We used the pure genetic algorithm option with R function `genoud` with disabled derivative information (Mebane, Jr. and Sekhon, 2011). `genoud` uses

a generation based optimisation, modifying a population of trial solutions so each generation is expected to be better than its predecessor, measured against the function to be optimised. Using derivatives is not mandatory, but in some cases might speed up the convergence on local optima. However, optimisation methods that rely on derivatives of the objective function may be unable to find optimum values when dealing with nonlinear functions and integer variables due to non-concavity and discontinuities (Sekhon and Mebane, Jr., 1998).

We implemented the same algorithm for three different functions to be optimised.

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 de-seasonalised aggregated demand data.

4.4. Minimisation of SD on seasonal signal

From $\mathbf{S}_{Cus \times 168}$ I get the partition $\mathbf{S}'_{SelectCus \times 24}$ where *SelectCus* are only the rows of the customers assigned to that group. I then calculate the stan-

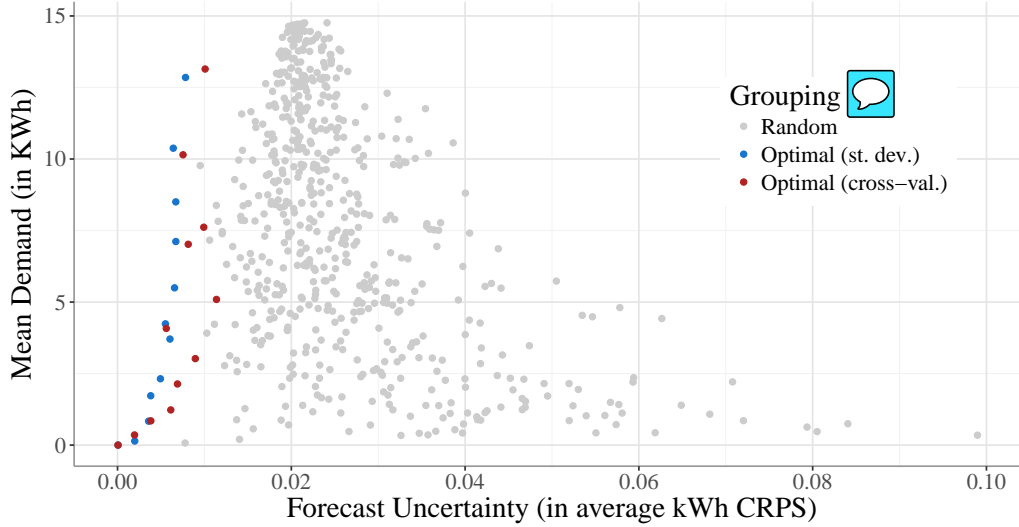


Figure 5: Optimal grouping of customers

dard deviation of each column of \mathbf{S}' and then the mean of all results. The difference between the models are the combination of \mathbf{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 have to talk to Fotios about his questions - email

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

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...

4.7. Discussion

relaxation of binary variables combination of SDEV and SEAS

5. Conclusion

References

- Al-Hamadi, H., Soliman, S., 2004. Short-term electric load forecasting based on kalman filtering algorithm with moving window weather and load model. *Electric Power Systems Research* 68 (1), 47 – 59.
URL <http://www.sciencedirect.com/science/article/pii/S0378779603001500>
- Arora, S., Taylor, J. W., 2016. Forecasting electricity smart meter data using conditional kernel density estimation. *Omega* 59, Part A, 47 – 59, *business Analytics*.
URL <http://www.sciencedirect.com/science/article/pii/S0305048314001546>

- Bonami, P., Lejeune, M. A., 2009. An exact solution approach for portfolio optimization problems under stochastic and integer constraints. *Operations Research* 57 (3), 650–670.
URL <https://doi.org/10.1287/opre.1080.0599>
- Chang, T.-J., Meade, N., Beasley, J., Sharaiha, Y., 2000. Heuristics for cardinality constrained portfolio optimisation. *Computers & Operations Research* 27 (13), 1271 – 1302.
URL <http://www.sciencedirect.com/science/article/pii/S030505489900074X>
- Cleveland, R. B., Cleveland, W. S., Terpenning, I., 1990. Stl: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics* 6 (1), 3.
- Feinberg, E. A., Genethliou, D., 2005. Load forecasting. In: *Applied mathematics for restructured electric power systems*. Springer, pp. 269–285.
- Garcia, R. C., Contreras, J., Van Akkeren, M., Garcia, J. B. C., 2005. A garch forecasting model to predict day-ahead electricity prices. *IEEE transactions on power systems* 20 (2), 867–874.
- Gneiting, T., Balabdaoui, F., Raftery, A. E., 2007. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 69 (2), 243–268.
URL <http://dx.doi.org/10.1111/j.1467-9868.2007.00587.x>
- Gooijer, J. G. D., Hyndman, R. J., 2006. 25 years of time series forecasting. *International Journal of Forecasting* 22 (3), 443 – 473, twenty five years

of forecasting.

URL <http://www.sciencedirect.com/science/article/pii/S0169207006000021>

Hain, J., Ault, G., Galloway, S., Cruden, A., McDonald, J., 2005. Additional renewable energy growth through small-scale community orientated energy policies. *Energy Policy* 33 (9), 1199 – 1212.

URL <http://www.sciencedirect.com/science/article/pii/S0301421503003562>

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

URL <http://www.sciencedirect.com/science/article/pii/S0169207015001508>

Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., Hyndman, R. J., 2016. Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. *International Journal of Forecasting* 32 (3), 896 – 913.

URL <http://www.sciencedirect.com/science/article/pii/S0169207016000133>

Jeon, J., Taylor, J. W., 2016. Short-term density forecasting of wave energy using arma-garch models and kernel density estimation. *International Journal of Forecasting* 32 (3), 991 – 1004.

URL <http://www.sciencedirect.com/science/article/pii/S0169207015001387>

Mebane, Jr., W. R., Sekhon, J. S., 2011. Genetic optimization using derivatives: The rgenoud package for R. *Journal of Statistical Software* 42 (11), 1–26.

URL <http://www.jstatsoft.org/v42/i11/>

- Pouttu, A., Haapola, J., Ahokangas, P., Xu, Y., Kopsakangas-Savolainen, M., Porras, E., Matamoros, J., Kalalas, C., Alonso-Zarate, J., Gallego, F. D., et al., 2017. P2p model for distributed energy trading, grid control and ict for local smart grids. In: Networks and Communications (EuCNC), 2017 European Conference on. IEEE, pp. 1–6.
- R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL <http://www.R-project.org/>
- Ramanathan, R., Engle, R., Granger, C. W., Vahid-Araghi, F., Brace, C., 1997. Short-run forecasts of electricity loads and peaks. International Journal of Forecasting 13 (2), 161 – 174.
URL <http://www.sciencedirect.com/science/article/pii/S0169207097000150>
- Schwarz, G., 03 1978. Estimating the dimension of a model. Ann. Statist. 6 (2), 461–464.
URL <http://dx.doi.org/10.1214/aos/1176344136>
- Sekhon, J. S., Mebane, Jr., W. R., 1998. Genetic optimization using derivatives: Theory and application to nonlinear models. Political Analysis 7, 189–213.
- Silverman, B. W., 1986. Density estimation for statistics and data analysis. Vol. 26. CRC press.
- Soliman, S., Persaud, S., El-Nagar, K., El-Hawary, M., 1997. Application of least absolute value parameter estimation based on linear programming to

- short-term load forecasting. *International Journal of Electrical Power & Energy Systems* 19 (3), 209–216.
- Suganthi, L., Samuel, A. A., 2012. Energy models for demand forecasting—a review. *Renewable and Sustainable Energy Reviews* 16 (2), 1223 – 1240.
URL <http://www.sciencedirect.com/science/article/pii/S1364032111004242>
- Tay, F. E., Cao, L., 2001. Application of support vector machines in financial time series forecasting. *Omega* 29 (4), 309 – 317.
URL <http://www.sciencedirect.com/science/article/pii/S0305048301000263>
- Taylor, J. W., Jeon, J., 2015. Forecasting wind power quantiles using conditional kernel estimation. *Renewable Energy* 80, 370 – 379.
URL <http://www.sciencedirect.com/science/article/pii/S0960148115001123>
- Wheelwright, S. C., Makridakis, S. G., et al., 1973. *Forecasting methods for management*. Wiley.