Forecast-based portfolio optimisation of electricity demand for P2P markets

Estêvão Ferreira Sêco de Alvarenga[1,*], Jooyoung Jeon[1,*], Ran Li[2,*], Fotios Petropoulos[1,*], 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. 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, the number of minimum customers required in P2P when the aggregation is constructed in a random manner. The research further suggests methods for 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 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 24h ahead. This approach is useful for peer to peer (P2P) energy trading that aims to utilise surplus in renewable energy generation locally. Our research indicate that a simple method based

[*]efsa@bath.edu

[1]University of Bath School of Management

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

on the standard deviation calculation using in-sample data performs as good as more complicated and time consuming approaches.
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. Electricity utilities have always forecasted 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).

In a near future, 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.

The fast pace adoption of breakthrough distributed generation technologies, electric vehicles (EV), and energy storage (Pouttu et al., 2017), coupled with empowered markets responding to economical signals, is likely to increase load uncertainty. Pouttu et al. (2017) suggests that "Aggregators" could support the market by reducing their inherited uncertainty. These actors would be responsible for procuring aggregation of demand, combining multiple short-duration consumer loads, and trading on organised energy market auctions. To enable this process, bespoke platforms that allow the trading optimisation for energy demand and generation are necessary. These platforms would be able to monitor, predict, schedule and make decisions regarding local energy patterns in real time (Mocanu et al., 2016).

Our research extends the risk-return relationship, from portfolio optimisation theory, for energy demand aggregation based on time-series forecasting. We hypothesise that a portfolio of customers, chosen through maximising the forecast accuracy, would be an ideal trading object from the energy demand perspective. This forecast-based portfolio optimisation platform would be able to support trade within the envisioned P2P energy marketplace as aggregators would be interested in matching this low error demand with distributed generation forecasts.

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). Having said that, wherever there is additional datasets available, regression based models could be used to enhance accuracy. This have been proved to enhance the prediction accuracy in noisy financial time series forecasting (Tay and Cao, 2001).

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

2. Literature review

2.1. P2P energy markets

In the UK, programmes supporting the 'community approach' to technology installation has supported the growth of multiple renewables energy technologies through a diverse of organisational and ownership models (Walker et al., 2010). As the deployment of distributed energy generation challenged the power system operation, they were simplified as load reduction assets at low penetration level. As their capacity rises, new roles are envisioned for the energy market.

The prosumers, a consumer that also generates and sell energy to the community, will be able to embrace more strategic opportunities towards active roles in the power market. Future prosumer markets can be envisaged as more or less structured, depending on social and policy choices. A fully decentralised market would enable peer-to-peer trading with little supervision. Additionally, prosumers could act as groups, and supply services to the power grid, such as frequency response or peak-shaving, supervised by

an aggregation node between the community and the system operator Moret and Pinson (2017).

(Pouttu et al., 2017) suggests that the Aggregator is a legal entity with commercial agreements with prosumers and able to trade in the energy market. This role would be responsible for aggregating load or generation of various resources to provide services with the potential to maintain power balance, quality and security. The Aggregator is a role that can be met by existing market actors or by a new separate business. This design would facilitate demand flexibility and distributed generation integration.

2.2. Energy demand forecasting

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). 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 (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).

Short to medium-term forecasting models have to take account the relationship between demand and weather. However, for lead times of 1-day-ahead there is little difference between performance compared to not using forecasted weather data. Non-weather models accuracy worsen as lead time increases, being dominated by more complex models (Taylor and Buizza, 2003).

When the interest is in lead times up to about six hours ahead, a univariate model is deemed to be sufficient. Whenever the availability of weather forecasts is patchy, univariate models are used for lead times larger than six hours ahead (Taylor et al., 2006). In the short run, the load is mainly influenced by meteorological real time conditions, seasonal effects (daily and weekly cycles, calendar holidays) and special events.

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). Literature provides a wide range of electricity demand forecasting methods, with probabilistic approaches being favoured due to non-linear and non-stationary demand profiles Mocanu et al. (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).

2.3. Cardinality constrained portfolio optimisation

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 portfolio of customers' demand does not behave like assets for a few reasons:

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

The cardinality constrained portfolio optimisation is an extension to the standard model that includes cardinality constraints that limit a portfolio

to a specified number of assets and impose limits on proportions on selected assets. Chang et al. (2000) present three algorithms based upon genetic algorithms, tabu search and simulated annealing for finding the cardinality constrained efficient frontier. The attraction of these algorithms are the effectively independency of the selected objective function, facilitating the assessment of different non-linear functions.

The properties of this portfolio optimisation extension, coupled with a solution that is valid for a given forecast lead time, would satisfy the unique characteristics of a portfolio of customers' demand, enabling the implementation of a forecast-based portfolio optimisation platform for P2P energy markets.

3. Experiment design

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 generation, a group of customers will be selected whose forecasted demand closely match the generation. Within the distribution network, this group could comprise from one to all customers.

The goal of our experiment is to look at the history demand for households in a community, and optimally group them for several different aggregated energy demands. This is done though defining a efficiency frontier on the risk-return relationship, based on cardinality constrained portfolio optimisation algorithms. The frontier (an example is given on Figure 1) relates the expected aggregated demand on the y-axis (return) and the forecast error on

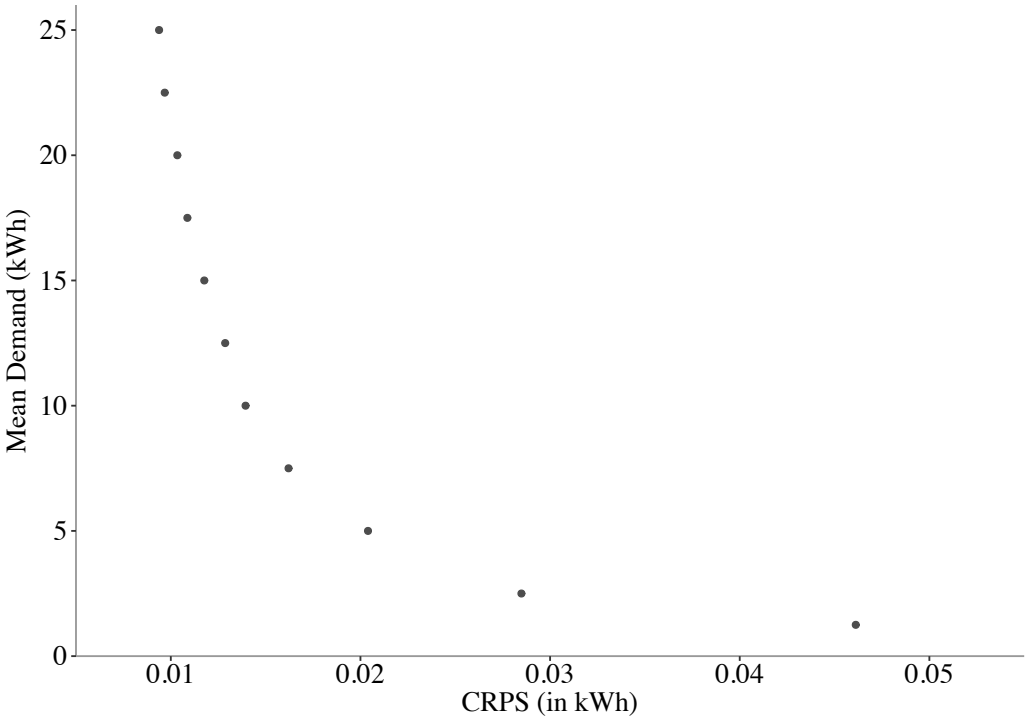

Figure 1: Example of risk-return frontier for forecast-based energy demand portfolios

the x-axis (risk).

maybe this figure should be on the first pages of the paper? Retrieving the expected future aggregated demand for a group of customers is a forecast based task. On the other hand, minimising the expected forecast error can be performed in many different ways, including forecast based methods. The Aggregator would use the frontier rendered by different methods to operate in the energy market minimising the risk of under- or over- bidding energy consumption, while trying to match supply offers from distributed energy generation.

The following pages describe this process. Starting from the seasonality

analysis (Section 3.1), the forecasting model and evaluation criteria (Section 3.2), and finally, the portfolio optimisation model (Section 3.3).

3.1. Seasonal analysis

Energy demand has natural cycles and the forecasting method we selected (described on Section 3.2) is not expected to produce good forecasts unless data has no trend. While some methods can handle seasonality, whenever the chosen one can not, it is considered a good strategy to analyse seasonal components, and remove them from the data, before running forecasts.

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). On the other hand, STL is limited to addictive decomposition, unless some form of data-transformation is implemented to obtain multiplicative decompositions.

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.

$$Y_t = T_t + S_t + R_t \tag{1}$$

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 $q \to \infty$ generates ordinary polynomial fits. STL consists on two procedures, updating seasonal and trend components, while computing robustness weights recursively. We implemented the STL seasonal analysis as the first step of our procedure. This was done using the R `stats` package (R Core Team, 2016).

3.2. Forecasting procedure

In this Section, we describe the forecasting procedure implemented. According to Taylor and Buizza (2003) there is no consensus as to the best approach to electricity demand forecasting. In this paper we implement an univariate autoregressive conditional heteroskedasticity (ARMA-GARCH) model supported by an unconditional kernel density estimation (KDE) model.

For each customer, in each time-step, we define an ARMA-GARCH model to produce forecasts, as described on Section 3.2.2. Whenever this model does not converge to an acceptable fit, we implement the KDE model, described on Section 3.2.1. The acceptable fit is defined via the adjusted Pearson chi-squared goodness-of-fit test, proposed by Vlaar and Palm (1993). Section 5.1 details the implementation and results for the dataset used on this paper.

Finally, on Section 3.2.3 we introduce the continuous ranked probability score (CRPS), the evaluation criteria of choice for the produced density forecasts.

3.2.1. 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 $\{Y_1, Y_2, ..., Y_n\}$. The unconditional KDE can be defined through Equation 2:

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

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:

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

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

3.2.2. 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 4a to 4c.

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

$$\sigma_t^2 = \delta_0 + \sum_{l=1}^{r} \delta_j \sigma_{t-l}^2 + \sum_{m=1}^{s} \gamma_k \varepsilon_{t-m}^2 \tag{4b}$$

$$\varepsilon_t = \sigma_t \eta_t \tag{4c}$$

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 distribution model.

We used the R package `rugarch` to build the ARMA-GARCH model (Ghalanos, 2014). The ARMA(p, q) order was defined via the lowest Bayesian

Information Criterion (BIC) value for the combination of possible (p, q) up to $(5, 5)$. BIC was described by Schwarz (1978) as an evaluation criteria for choosing the appropriate dimensionality of a model. The main difference between BIC and maximum likelihood or Akaike Information Criteria (AIC) is the application of a higher penalty for high order models. We chose to limit (p, q) up to $(5, 5)$ as we deemed this was a good trade-off between forecast accuracy and computing resources. The GARCH(r, s) order was set to (1,1) as this has been shown to produce accurate practical result for volatility estimation (Hansen and Lunde, 2001).

For the distribution selection, `rugarch` enables the use of normal, student-t, generalised error distribution and their skewed versions. Theodossiou (2015) investigated empirical distributions for finance data with the presence of skewness, kurtosis, conditional heteroskedasticity, asymmetric volatility, and other non-linear dependencies. This properties created biased models when applying normal distributions. A skewed version of the generalised error distribution was proven to provide better fit to the empirical distribution of the data. Based on this, and to accommodate a highly volatile dataset, we chose this distribution for the ARMA-GARCH model.

3.2.3. Evaluation criteria

With the forecasts in hands, we accordingly added back the deterministic seasonality removed previously, as described in Section 3.1. The trend-cycle component was extrapolated with the last value recorded in the in-sample data and added to the forecasts. 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 generalises the absolute error, to which it reduces in case of point forecasts, and assess probabilistic forecasts against deterministic observations. The CRPS is reported in the same unit as the observations.

The formal definition of this scoring method is described in Equation 5.

$$\text{CRPS} = \int_{-\infty}^{\infty} \text{BS}(y)dy \tag{5}$$

Where BS denotes the Brier Score (Brier, 1950) of probability forecasts, as in Equation 6.

$$\text{BS} = \frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)^2 \tag{6}$$

Where p_i is the density forecast and o_i is the observed value.

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.

3.3. Forecast-based Portfolio Optimisation

The cardinality constrained portfolio optimisation goal is to find, from possible groups of customers, smaller groups that would deliver high forecasting accuracy. The cardinality constrain is defined as a binary variable where each customer is included or not in a given portfolio.

We chose three different objective functions on our work, further detailed in this Section:

1. Forecast Validated (FV): minimising forecast error using the forecasting function on a validation dataset.
2. Seasonal Residual (SR): minimising standard deviation of the residual of a deseasonalised aggregated demand and;
3. Seasonal Similarity (SS): minimising the deviation of the seasonal signals of customers pertaining a portfolio;

All three implemented objective functions require a nonlinear function f_i. This function is applied to a matrix $\boldsymbol{A}$ containing the demand history (columns) per customer (rows) and a customer selection vector v. Considering the cardinality restrictions, v was defined as a binary vector with as many positions as the number of households (equation 7).

$$\underset{v_j \in \mathbb{N}[0,1]}{min} \big(f_i(A, v)\big) \tag{7}$$

To perform this minimisation, 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 (Chang et al., 2000). We used the pure genetic algorithm option with R function `genoud` disabling the derivative information option (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. This algorithm also allows for integer constraints to be implemented on the variable

space, outside the objective functions, which dramatically reduces computational time.

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).

Sections 3.3.1 to 3.3.3 describe the three proposed objective functions. Further, Section 5 present the results from different optimisation approaches, followed by discussions on Section 6.

3.3.1. Forecast Validated (FV) function

The FV method is the most straight forward of the three selection functions. The objective is the optimisation of the group selection based on the forecast accuracy. Specifically, the FV method divides $\boldsymbol{A}$ into a training matrix $\boldsymbol{A_a}$ and a test matrix $\boldsymbol{A_b}$. A density forecast g is then produced using the training matrix, while the evaluation occurs against the test matrix.

$$F_{fv}\big(g(\boldsymbol{A_a} \times v), \boldsymbol{A_b} \times v\big) \tag{8}$$

F_{fv} returns the forecast accuracy of the evaluation step. The goal of the optimiser is the minimisation of Equation 8, through changing v. The vector v that delivers the minimal value of F_{fv} is the customer selection that minimises the forecast error through the FV function.

3.3.2. Seasonal Residual (SR)

The SR method implements the function F_{sr} onto the undivided matrix $\boldsymbol{A}$ and the selection vector v, as in equation 9.

$$F_{sr}(\boldsymbol{A}, v) = R_t^{\boldsymbol{A}\times v} \tag{9}$$

$R_t^{\boldsymbol{A}\times v}$ is defined as the remainder of the seasonal decomposition (described on Section 3.1) onto the time-series generated after multiplying matrix $\boldsymbol{A}$ with the selection vector v, as in equation 10.

$$STL(\boldsymbol{A} \times v)_t = T_t^{\boldsymbol{A}\times v} + S_t^{\boldsymbol{A}\times v} + R_t^{\boldsymbol{A}\times v} \tag{10}$$

The goal of the optimiser is the minimisation of the SR function F_{sr} applied to $\boldsymbol{A}$, through changing v.

3.3.3. Seasonal Similarity (SS)

The SS method is a bi-objective optimisation using function F_{ss} applied onto the undivided matrix $\boldsymbol{A}$ through the selection vector v, using the objective ratio r. While both FV and SR use v to multiply $\boldsymbol{A}$ and generate a single time series to be used by the optimisation, SS uses v to create a new demand history matrix B that is a partition of $\boldsymbol{A}$, and includes only the rows of selected customers.

The function then performs a seasonal decomposition on each row i of B, following equation 11.

$$STL(\boldsymbol{B_t}) = \boldsymbol{T_t^B} + \boldsymbol{S_t^B} + \boldsymbol{R_t^B} \tag{11}$$

Matrix $\boldsymbol{S_t^B}$ contains the seasonal component for each customer, while matrix $\boldsymbol{R_t^B}$ contains the remainder of the seasonal decomposition for each customer. The bi-objective approach estimates the relative standard devi-

ation ($\widehat{RSD}$) per column and then averages the result per matrix, before applying the bi-objective ratio r.

$$F_{ss}(\boldsymbol{A}, v, r) = \frac{\sum_{j=1}^{n} \widehat{RSD}(\boldsymbol{S_t^B}_j)}{n}(r) + \frac{\sum_{j=1}^{n} \widehat{RSD}(\boldsymbol{R_t^B}_j)}{n}(1-r) \qquad (12)$$

Where RSD is given by the sample standard deviation s divided by the mean $\bar{x}$.

The first part of Equation 12 calculates the similarity between the seasonal components of selected customers. Lower $\widehat{RSD}(\boldsymbol{S_t^B})$ per column means minimal difference between seasonal signals per time-step. The second part of Equation 12 calculates the similarity between the remainder components of selected customers. Since it is expected that $\boldsymbol{R_t^B}$ has no trend, a second consequence is that lower $\widehat{RSD}(\boldsymbol{R_t^B})$ means customers with lower variability after removing the seasonal signal. Different ratio r can be used to change the weights between the objective functions.

The goal of the optimiser is the minimisation of Equation 12, through changing v for a given r.

4. The smart meter data

This Section introduces the data used in the experiment, 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. The daily and weekly seasonal cycles s_{01} and s_{02} are, respectively, 24 and 168. Figure 2 shows example of different

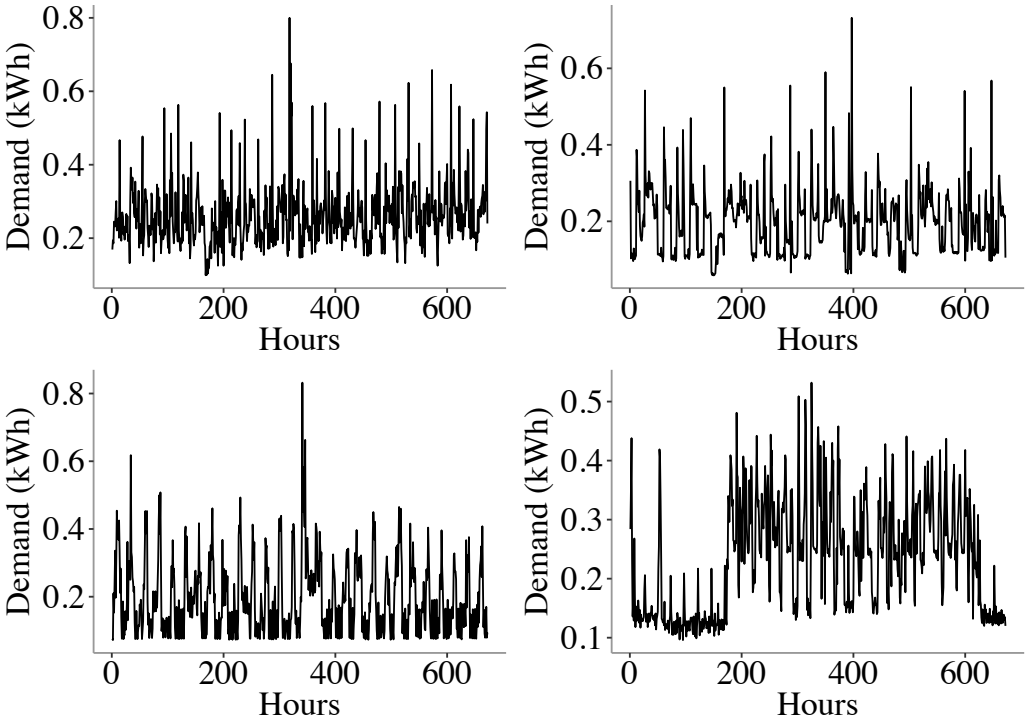

Figure 2: Example of electricity demand pattern for four different customers during a period of four weeks

behaviours during the same four week period for four different customers in February 2012. Additionally, Figure 3 exemplifies density plots for the same 24 hours for different customers.

To avoid some periods of missing observations, we did not use the first 12 weeks of data, which contained lots of empty data. If there still was periods with missing observations, we copied the value from previous week, from same weekday and hour. This approach will keep some seasonal pattern, while avoiding the use of out-sample data, used for accuracy calculations for testing models.

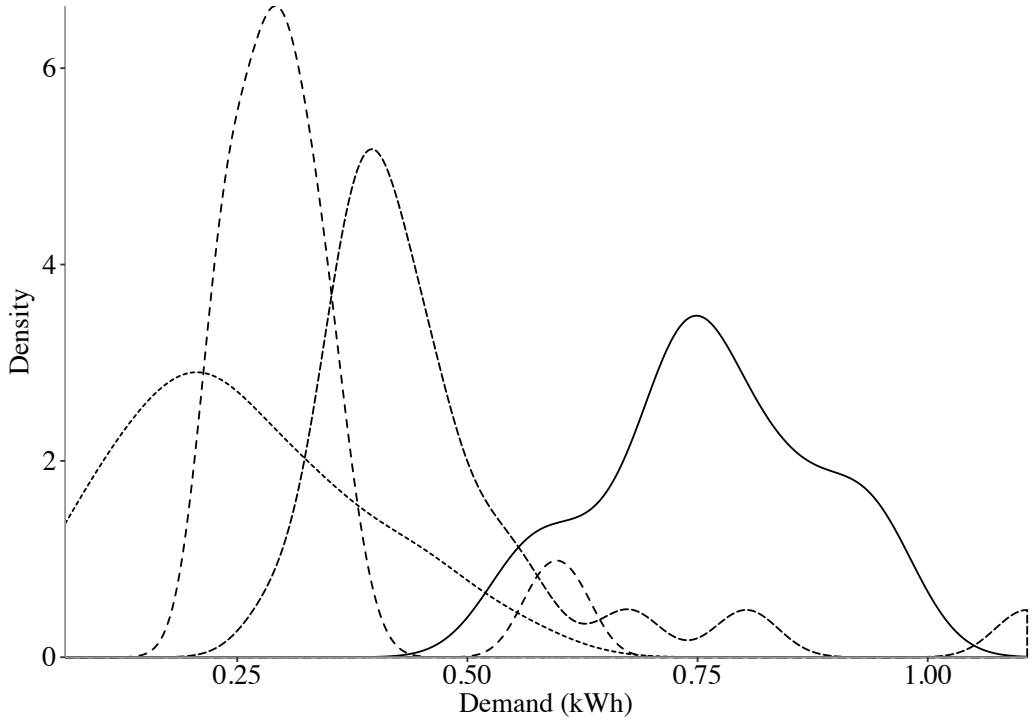

Figure 3: Example of electricity demand densities forecasted for four different individual customers, built using KDE with a window size of 24h, without prior seasonal analysis

Should I describe the Irish data here or on discussion, when I compare KO and IR.

Datasets containing 12 weeks were used for model fitting and forecast generation for up to the next 72 hour. We had access to a a very big high frequency dataset. Instead of using a rolling forecast approach, for each possible hour of available data, we summarised the data as follows: For the initial analysis on models and parameters, before any forecast results for the paper were collected, we used forecasting starting points 37 hours apart of each other. This allowed for the forecasts to start in different hours inside weeks. For data collected for results and discussion, we used sections of the dataset 113 hours apart of each other. This method would enable us to use data from several years for the initial analysis without disrupting the blindness of the datasets for forecasting. Also, it allowed us to research a big space of the original dataset, and not have a high bias toward certain week days or times of day, while restricting computational time to a feasible cost. not sure if I should keep this paragraph

5. Results

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 200 households. Figure 4 exemplifies how aggregating the demand of several customer reduces the instability of demand.

Each of these groups had their demand forecasted and evaluated through the procedure described on Section 3.2. In order to maintain proportional

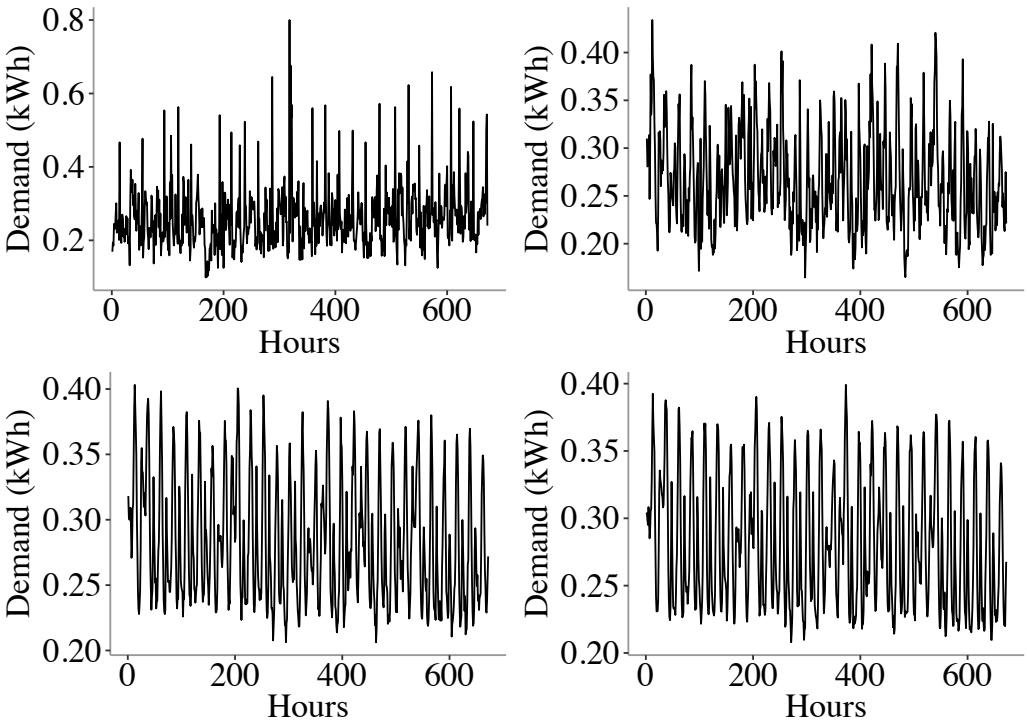

Figure 4: Example of demand for one customer (top-left), and aggregated demand for ten (top-right), one hundred (bottom-left) and two hundred (bottom-right) customers during a period of four weeks

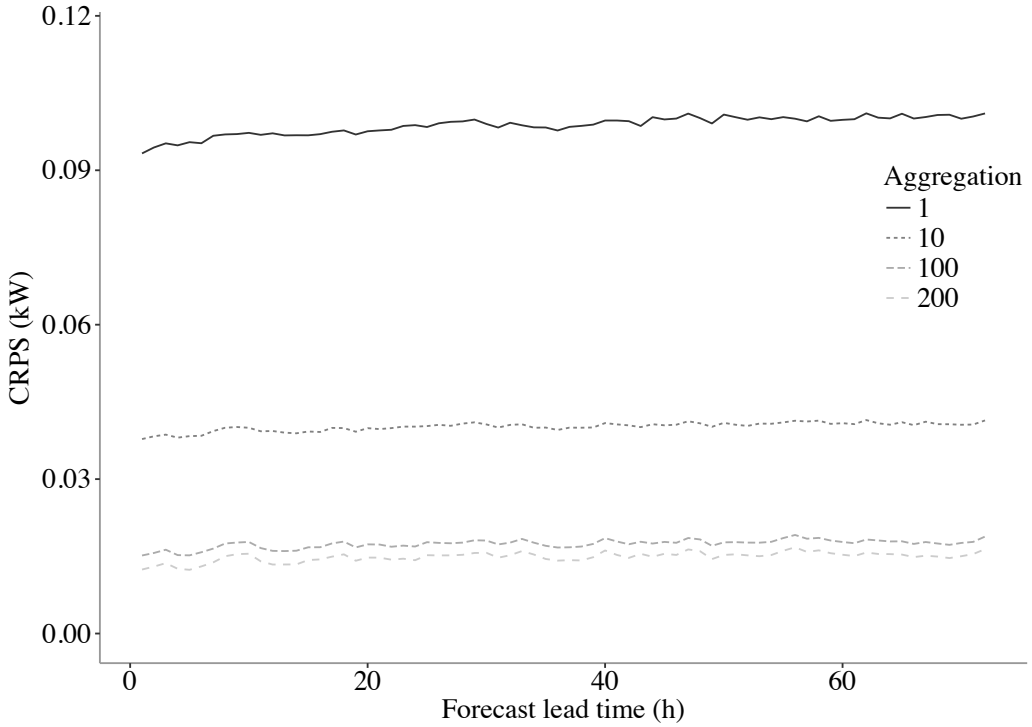

Figure 5: Forecast uncertainty for different sizes of random groups of households

comparability, the aggregated demand value of each group is the sum of its households' demand divided by the number of households.

Figure 5 represents the forecast accuracy versus the forecast lead time. As expected, forecast error increases (CRPS value) as the lead time increases. A strong converging pattern can be seen The inaccuracy has a major decrease from individual households to groups of 10, while decreases marginally when comparing groups of 100 and 200 households. This suggests that only increasing the number of individuals per group would not be a optimal solution for reducing forecast uncertainty. Nevertheless, a larger aggregation is always expected to generate forecasts with lower errors.

This groups created following a random fashion were used every time a portfolio optimisation model was run. The goal is to visualise where the proposed efficient frontier would seat, comparing to the basic random aggregation.

Since the optimisation depends on a calibrated forecast function, Section 5.1 describes how the fit convergence for ARMA-GARCH models were defined. This parameter is responsible for switching from an ARMA-GARCH-based forecast that is likely to perform bad to a KDE-based forecast, without peeking into the out-sample dataset. Sections 5.2 and 5.3 then introduces how the genetic optimisation and the functions defined before were set-up for the portfolio optimisation problem in hand. Finally, section 6.3 presents the results for this entire process for different functions and different forecast lead-times.

5.1. Fit convergence for ARMA-GARCH models

For every fitted ARMA-GARCH model we performed an adjusted Pearson chi-squared goodness-of-fit test, proposed by Vlaar and Palm (1993), through the `rugarch` package. The produced goodness-of-fit δ was compared against a global threshold previously defined.

To define the best global threshold δ to be used for the forecast model, we implemented five different configurations. One based on KDE without any ARMA-GARCH coding, so we can compare results against a benchmark, and four ARMA-GARCH based models, as respectively defined in Sections 3.2.1 and 3.2.2. The difference between each ARMA-GARCH models was the global threshold varying from $\delta = 0$ to $\delta = 0.2$.

We ran a simulation containing 300 customers at 250 different sections

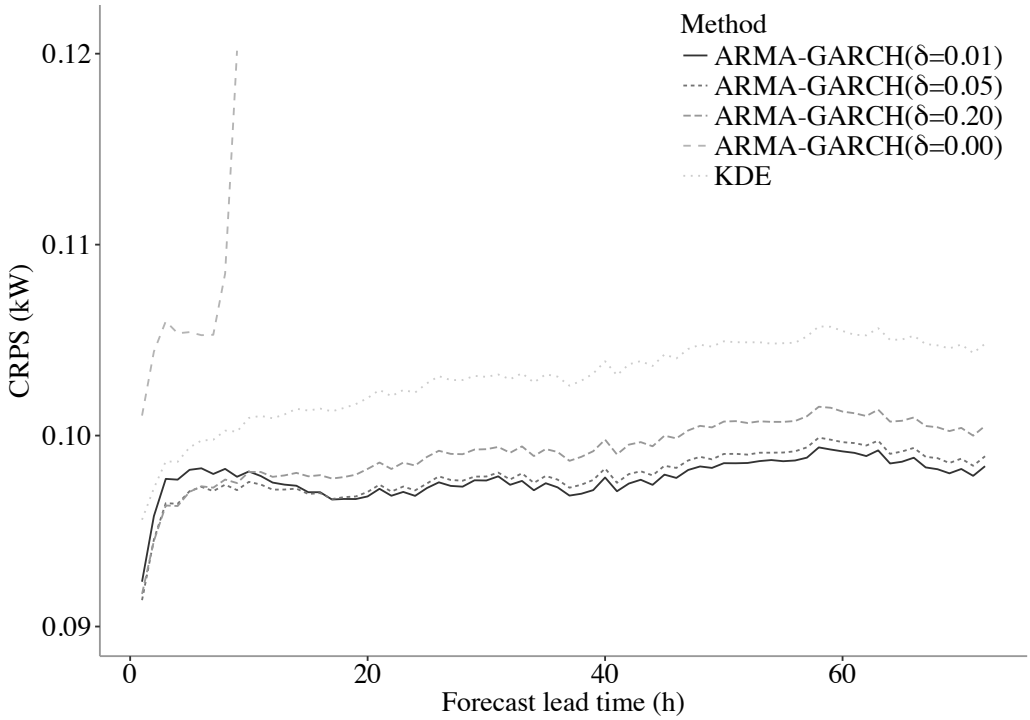

Figure 6: Goodness-of-fit selection

from the data, from September 2012 to October 2013. We produced individual household's post-sample forecasts from 1 to 72 hours ahead, after each in-sample period of 12 weeks, and averaged the CRPS values. Figure 6 compares the forecast accuracy for different δ values.

For this particular dataset, when forecasting for lead times smaller than 12 hours, the ARMA-GARCH ($\delta = 0.05$) model is the best performer. This model starts to underperform, compared to ARMA-GARCH ($\delta = 0.01$), at some point between lead times of 12 and 18 hours.

5.2. Genetic algorithm optimisation

The genetic algorithm was set to stop after 100 generations or after 10 consecutive generations without enhancement in the function value, whatever happened first. For all methods, we used 12 weeks as the in-sample period (matrix $\boldsymbol{A}$, containing 100 customers as rows, and 2016 time-steps as columns). Instead of running the optimisation to find one unique optimal group, we divided the expected demand axis into 16 bins. This method was chosen to enable the construction of a stepwise efficiency frontier for each optimisation function. Practically this means that, for any chosen function, instead of running a single optimisation run on $\boldsymbol{A}$ for the full spectrum of expected demand, we run 16 optimisations with different and non-superpositioned expected demand limits. The process created groups whose aggregated demand would be between given minimum and maximum values. This number of divisions was selected as they enabled the construction of a piecewise frontier while allowing the entire process to be executed in a feasible computational time.

5.3. Optimisation functions set-up

The SR function was implemented as described on Section 3.3.2. For FV and SS functions, the following paragraphs describe additional configurations that were made to allow the best implementation of those functions for the aforementioned dataset.

Specifically for the FV function, we divided $\boldsymbol{A}$ into $n = 16$ partitions to enable multiple runs in a rolling forecast fashion. This allowed the forecast evaluation of several positions in the in-sample dataset against the same selection vector v. This number of divisions rendered a good trade-off between

Table 1: Results for different in-sample period oversaw by FV optimisation

In-sample period	**Average CRPS**
1 day	0.006921313
7 days	0.006198084

optimisation result, overfitting, and computational time.

F_{fv} was applied to each training matrix $\boldsymbol{A}_{n,a}$ and evaluated against the test matrix $\boldsymbol{A}_{n,b}$, using the same selection vector v. The average of these evaluations was returned as the FV function result, and as described in Section 3.3.1, v was chosen in order to minimise this result. An alternative approach, that selected only the last day instead of the last week of the in-sample period, was implemented as well, rendering less accurate forecasts through the chosen selection vector v (see Table 1).

For the SS function, three groups of options were explored on the process described on Section 3.3.3.

1. prior execution of a normalisation routine on component S_t^B before comparing the similarity of the seasonal behaviour;
2. either focusing the similarity comparison on the selected forecast lead time of S_t^B or on the entire seasonal component;
3. changing the weights for the bi-objective function

The different results on forecast accuracy for each combination of options 1 and 2 are presented on Table 2. The non-normalisation of seasonal component while considering the entire in-sample dataset, instead of focusing on the forecast lead time, rendered the best result for Korean data.

Table 2: Results for SS function running with different configuration on prior normalisation routine and forecast lead-time focus for similarity evaluation

Configuration	**Average CRPS**
Non-normalised and focused	0.008493832
Normalised and focused	0.008264960
Non-normalised and non-focused	0.006933789
Normalised and non-focused	0.008999768

Table 3: Results for SS function running with different configuration on weights given to bi-objective function

Configuration	**Average CRPS**
Seasonal component given weight 1.0	0.007203195
Seasonal component given weight 0.8	0.008410930
Seasonal component given weight 0.5	0.007606566

Having selected the first two options, the different weights were assessed and Table 3 summarises the findings. For the same dataset, giving a weight $r = 0.1$ to the seasonal component, while ignoring the remainder similarity on the bi-objective function, returned the best accuracy.

5.4. Portfolio optimisation results for Korean data

Figures 7 and 8 present the portfolio optimisation for two different time-ahead forecasts. This results are the average values from 11 different simulations, each containing 100 customers from the Korean dataset. Table 4 summarises the results, averaging the accuracy value throughout the en-

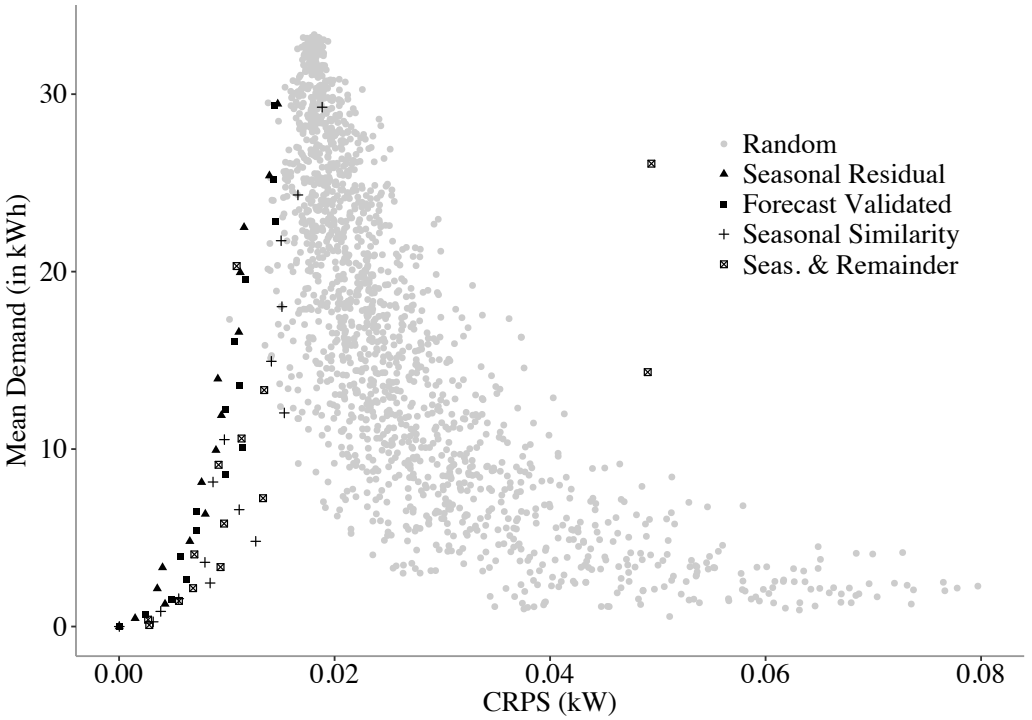

Figure 7: Portfolio optimisation using a 4h-ahead forecast for Korean customers

ergy demand spectrum for all functions. All three optimisation functions implemented deliver good result at selecting groups that would reduce the uncertainty for given mean demand ranges. As the mean demand increases, a simple random aggregation of customers are able to reduce most of the uncertainty, leaving a limited space for optimisation (upper part of the figures). Nevertheless the portfolio optimisation returns the best available groups (or very close to it) most of the time, specially for the SS and FV functions. On both forecast lead-times the SS functions perform poorly compared to the alternatives. Also, one can see that a longer forecast lead-time render a worse accuracy for random groups of customers.

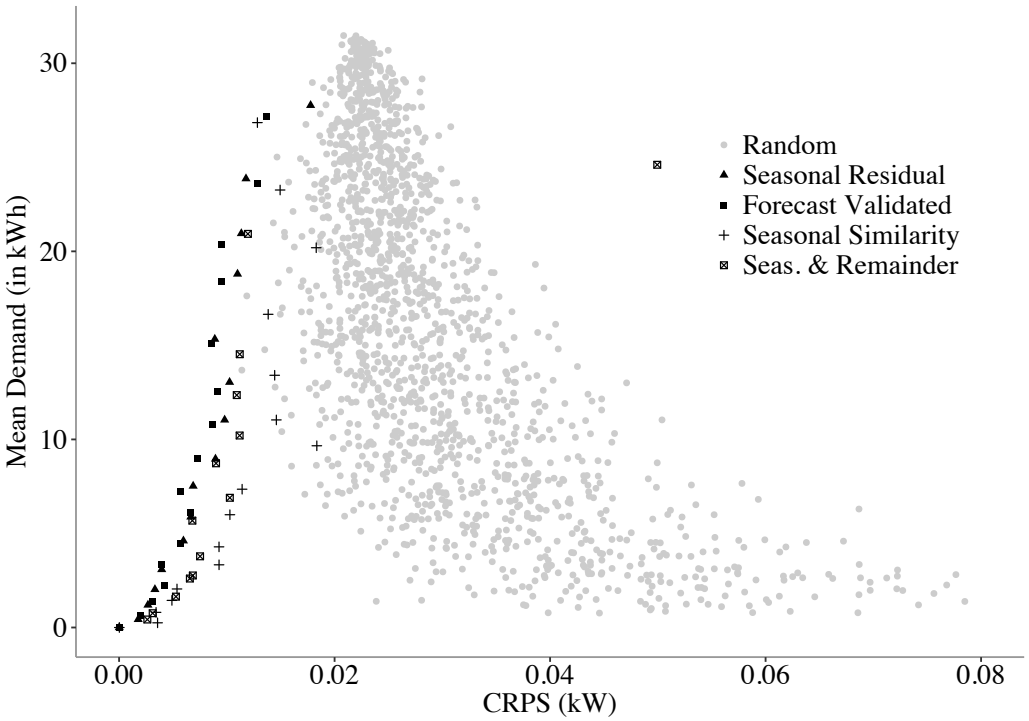

Figure 8: Portfolio optimisation using a 12h-ahead forecast for Korean customers

Table 4: Average portfolio optimisation results for different functions and lead-times, using Korean data

Optimisation Function	**Average CRPS 4h-ahead**	**Average CRPS 12h-ahead**
Random	0.026942760	0.029939163
Seasonal Residual	0.007866620	0.007564032
Forecast Validated	0.008857013	0.006906969
Seasonal Similarity	0.010399700	0.010303772

6. Discussion

6.1. Comparison between SR and FV functions

The Seasonal Residual (SR) function assumes the same group regardless of the forecast lead time, selecting customers with a high likelihood of minimal standard deviation from a deterministic seasonal decomposition, based on history demand. On the other hand, the Forecast Validated (FV) approach allows us to find the optimal composition of households on a certain forecast lead time of our interest. In respect to that, a group created for a $t_{f1} = 4$ forecast might not be the best grouping for a $t_{f2} = 24$.

Another characteristic of the FV method is that it allows groups of customers with high variance aggregated remainder signal to be included, as long as their forecast is stable throughout the training dataset. This can be seen as an advantage, as it increases the opportunities for grouping, and as an disadvantage as well, as it enables groups with unforeseen inaccurate forecasts to be created. Another drawback of this functions is the presence of overfitting as groups can be created perfectly when looking at the training dataset, while their performance on the test dataset does not hold. In someway, the advantages and disadvantages seem to compensate each other, as the FV function performs slightly better for a longer lead-time, and slightly worse for the short lead-time.

Computationally speaking, any signal analysis method (SR or SS) consumes significantly lower time to run as they do not require many forecasting runs while grouping customers. The FV function needs thousands of forecast runs while grouping, and this can get very time intensive. On our experiments, with our subset of data and a 2016 8-core i7 processor, FV optimi-

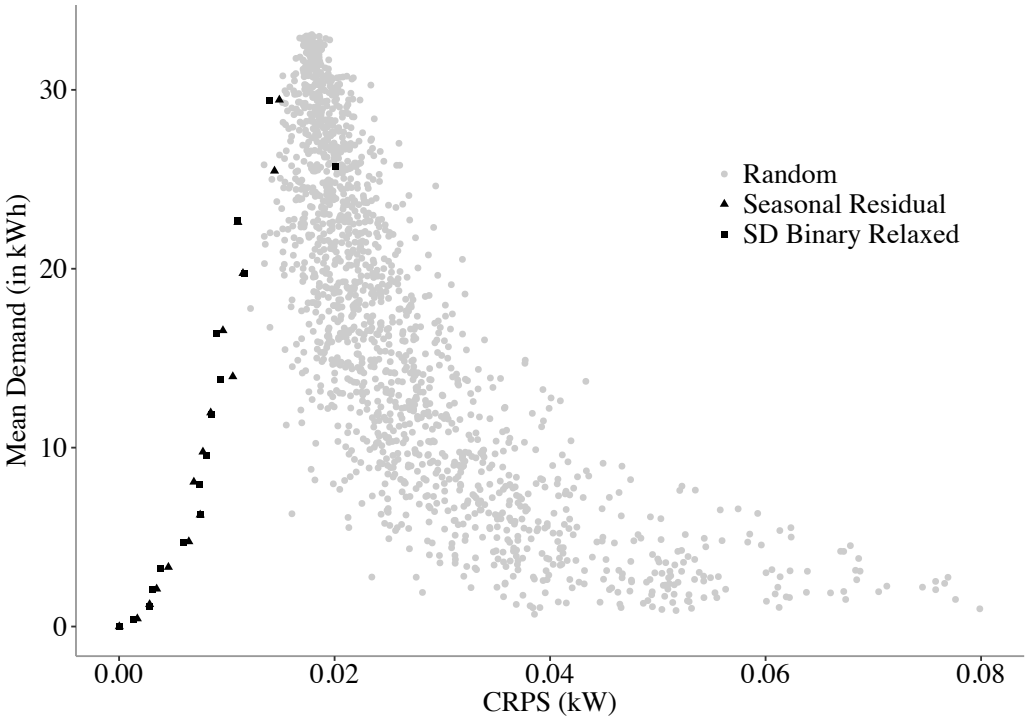

Figure 9: Portfolio optimisation relaxing the binary constraint on the 4h-ahead forecast for Korean customers

sation costs 27h to build an entire frontier for a single simulation, while the SD function runs ins less than 1h.

6.2. Relaxing the binary constraint

In order to assess if a relaxation of the binary constraint would produce different results, we chose to implement a SR function allowing fractions of each household to be selected for grouping (figure 10). It can be seen as there is not much advantage in allowing this selective grouping, with both relaxed and unrelaxed versions of SR performing similarly.

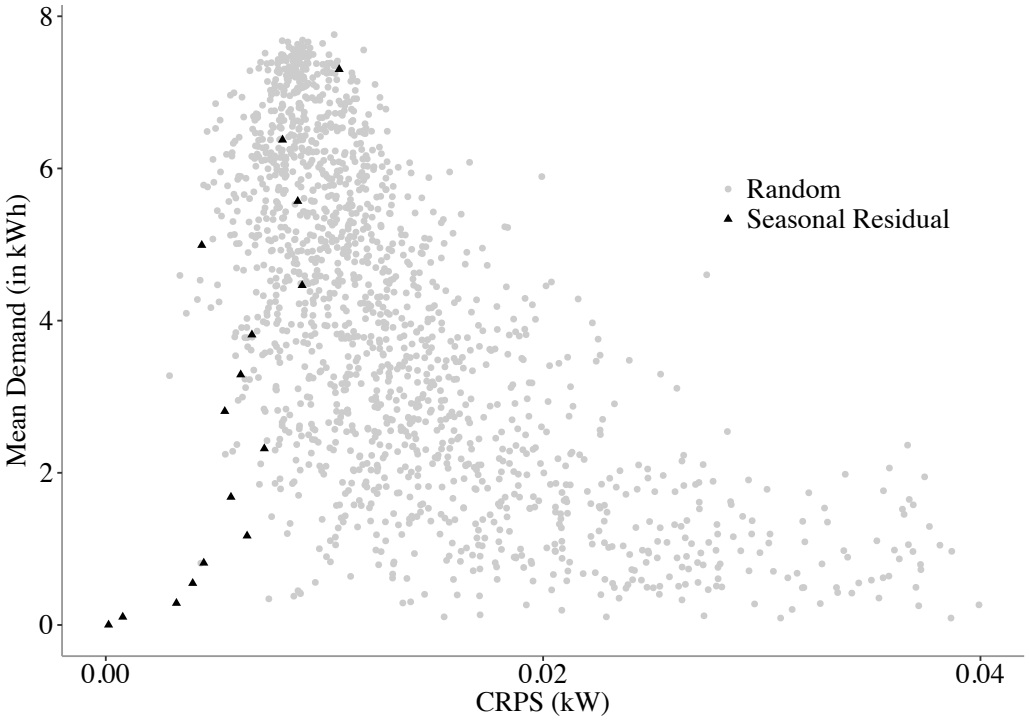

Figure 10: Portfolio optimisation using a 4h-ahead forecast for Irish customers

6.3. Portfolio optimisation with the Irish data

Comparison of the two data sets

In addition, we run the same SR function using the Irish dataset. The results are still good to generate groups on the lower half of the aggregated demand, but not overcoming as good as the number of customers needed for grouping increases. Specially for very large aggregations, this method produces groups that are worse than the average of a random grouping method.

7. 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

Brier, G. W., 1950. Verification of forecasts expressed in terms of probability. Monthey Weather Review 78 (1), 1–3.

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.

Ghalanos, A., 2014. rugarch: Univariate GARCH models. R package version 1.3-5.

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

Hansen, P. R., Lunde, A., 2001. A comparison of volatility models: Does anything beat a garch (1, 1). Unpublished manuscript. Department of Economics, Brown University.

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/

Mocanu, E., Nguyen, P. H., Gibescu, M., Larsen, E. M., Pinson, P., 2016. Demand forecasting at low aggregation levels using factored conditional restricted boltzmann machine. In: Power Systems Computation Conference (PSCC), 2016. IEEE, pp. 1–7.

Moret, F., Pinson, P., 2017. Energy collectives: a community and fairness based approach to future electricity markets. Energy Collective project.

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., Buizza, R., 2003. Using weather ensemble predictions in electricity demand forecasting. International Journal of Forecasting 19 (1), 57 – 70.
URL http://www.sciencedirect.com/science/article/pii/S0169207001001236

Taylor, J. W., de Menezes, L. M., McSharry, P. E., 2006. A comparison of univariate methods for forecasting electricity demand up to a day ahead. International Journal of Forecasting 22 (1), 1 – 16.
URL http://www.sciencedirect.com/science/article/pii/S0169207005000907

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

Theodossiou, P., 2015. Skewed generalized error distribution of financial assets and option pricing. Unpublished manuscript. Social Science Research Network (SSRN).
URL https://ssrn.com/abstract=219679

Vlaar, P. J., Palm, F. C., 1993. The message in weekly exchange rates in the european monetary system: Mean reversion, conditional heteroscedasticity, and jumps. Journal of Business & Economic Statistics 11 (3), 351–360.
URL http://amstat.tandfonline.com/doi/abs/10.1080/07350015.1993.10509963

Walker, G., Devine-Wright, P., Hunter, S., High, H., Evans, B., 2010. Trust and community: Exploring the meanings, contexts and dynamics of community renewable energy. Energy Policy 38 (6), 2655 – 2663, the Role of Trust in Managing Uncertainties in the Transition to a Sustainable Energy Economy, Special Section with Regular Papers.
URL http://www.sciencedirect.com/science/article/pii/S0301421509003541