# Probabilistic load forecasting for the low voltage network: Forecast fusion and daily peaks

Ciaran Gilbert [a], Jethro Browell [b],*, Bruce Stephen [a]

[a] *Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, G1 1XQ, UK*
[b] *School of Mathematics and Statistics, University of Glasgow, G12 8TA, UK*

## ARTICLE INFO

## ABSTRACT

Short-term forecasts of energy consumption are invaluable for the operation of energy systems, including low voltage electricity networks. However, network loads are challenging to predict when highly desegregated to small numbers of customers, which may be dominated by individual behaviours rather than the smooth profiles associated with aggregate consumption. Furthermore, distribution networks are challenged almost entirely by peak loads, and tasks such as scheduling storage and/or demand flexibility maybe be driven by predicted peak demand, a feature that is often poorly characterised by general-purpose forecasting methods. Here we propose an approach to predict the timing and level of daily peak demand, and a data fusion procedure for combining conventional and peak forecasts to produce a general-purpose probabilistic forecast with improved performance during peaks. The proposed approach is demonstrated using real smart meter data and a hypothetical low voltage network hierarchy comprising feeders, secondary and primary substations. Fusing state-of-the-art probabilistic load forecasts with peak forecasts is found to improve performance overall, particularly at smart-meter and feeder levels and during peak hours, where improvement in terms of CRPS exceeds 10%.

## 1. Introduction

Distribution networks are in a state of transition, with their original remit of 'fit and forget' now supplanted by the potential of massive increase in utilisation from electrified transport and heat, and two-way power flow from embedded renewable generation. This has led to the emerging necessity of the Distribution System Operator (DSO), who holds responsibility for balancing power flows under the transmission network [1]. While balancing has been commonplace at transmission-level for decades and the requisite forecasting and dispatch capabilities well understood, there is not a direct translation from transmission to distribution. Going down the voltage levels in power networks makes individual, low diversity, demand behaviours less predictable and hence unsuited to the methods used at transmission and regional levels. Short-term forecasts of load at all levels of the distribution network will be essential to coordinate flexibility services from distributed energy resources.

Load forecasting on the transmission network is a highly active area of research, and has been a mature technology for decades. Research in recent years has been focused on probabilistic forecasts, which communicate the uncertainty associated with a forecast to end-users. There is a growing appetite for such forecasts in industry, which are now used by both Transmission System Operators (TSOs) and energy traders in operational decision making. Probabilistic load forecasting is extensively reviewed in [2], where the authors highlight the emergence of household level forecasting in the context of hierarchical modelling; an opportunity provided by newly available smart meter datasets.

Forecasting at Low Voltage (LV) levels poses a different challenge to the conventional load forecasting problem at the transmission level. As electricity is aggregated group behaviours emerge which tend to change slowly and are therefore relatively predictable. Disaggregated demand at the household level is much more changeable and influenced by individual behaviours and processes, as shown in Fig. 1. The effect of the signal to noise ratio at the various voltage levels is discussed in much of the following literature where it is suggested that new approaches to forecasting are required and that should be developed with end-use in mind.

Apart from the challenge of the lower signal to noise ratio at the household level, there are also challenges relating to the large number of nodes in LV networks where forecasts may be required, limited coverage of monitoring, data quality, and data privacy. These constraints affect the applicable methodologies for forecasting; for instance the models must be computationally

* Corresponding author.
   *E-mail address:* jethro.browell@glasgow.ac.uk (J. Browell).

efficient and the input features may not be location specific. Challenges and opportunities for low voltage forecasts are discussed in an extensive review of the literature [3,4], where the authors outline recommendations for future research; such as the need for probabilistic forecasting, handling limited observability, robust forecast verification, and avoiding widespread single-source data-bias in research projects.

The smart meter roll-out in Great Britain and around the globe presents new opportunities in household load forecasting. This area has received the most attention in the literature around LV forecasting, which has mainly focused on deterministic forecasting [5,6]. The high penalisation of phase or 'timing' errors by traditional point-wise deterministic metrics like mean absolute error is highlighted in [7], where a new evaluation measure based on temporal permutation is proposed. The problem highlighted is that typical evaluation metrics tend to reward a smoother forecast on average compared to a forecast that may better represent the underlying process, that misses the precise timing of a sharp increase in demand. Therefore, the importance of predicting peaks in household (and LV) electricity demand has been discussed extensively in the literature [4].

The volatility of household demand necessitates a probabilistic approach to forecasting. Univariate probabilistic forecasts are typically communicated as full density forecasts, which are the most flexible for use in decision making, or in the form of multiple discrete quantiles at various probability levels. Interval forecasts with a specific coverage probability are also common [8]. In [9] density forecasts are obtained using Kernel Density Estimation (KDE), but conditional on information such as time-of-day, with a boundary correction applied to account for the positive nature of demand. Similarly, beta kernels are used in [10] to address the same problem, with a focus on building a scalable forecasting algorithm. Full density forecasts are generated as a benchmark model in [11], where the conditional density is assumed to be Gaussian with variance conditional by time-of-day; this is compared to non-parametric forecasts produced using an LSTM network for quantile regression, which outperforms the conditional density approach for the quantiles considered. A quantile regression approach based on boosting with additive models is demonstrated in [12], where the additive models are flexible and benefit from automatic feature selection by nature of the component-wise boosting procedure. Importantly, the quantile forecasts at the smart meter level are shown to be more skilful than an advanced parametric approach based on the Gaussian distribution. Bernstein polynomials have also been proposed for producing non-parametric density forecasts in [13] which show improvement over Gaussian and Gaussian mixture density forecasts. Finally, multivariate forecasts are generated for a hierarchy of smart meters in [14], where a coherency constraint is placed on the samples of the multivariate distribution, i.e. lower levels must sum to higher levels of the hierarchy. This literature may give the impression that only non-parametric densities are suitable for household load forecasting, however, there have been no studies examining alternatives to the Gaussian distribution until the present work.

Some of the works already discussed (e.g. [12,14]) also forecast low- and medium-voltage levels, i.e. feeder and substation load. However, the networks are hypothetical, in that they are generated from aggregated smart meter data. As discussed in [4], this necessarily excludes important elements on the distribution network such as commercial loads, street lighting, embedded generation, electrical losses, dependencies between customers, etc. However, there is a shortage of open-source data sets that contain a satisfactory amount of nodes due to a lack of widespread monitoring. An alternative option is to generate a synthetic LV dataset from smart meter data and basic information on the local
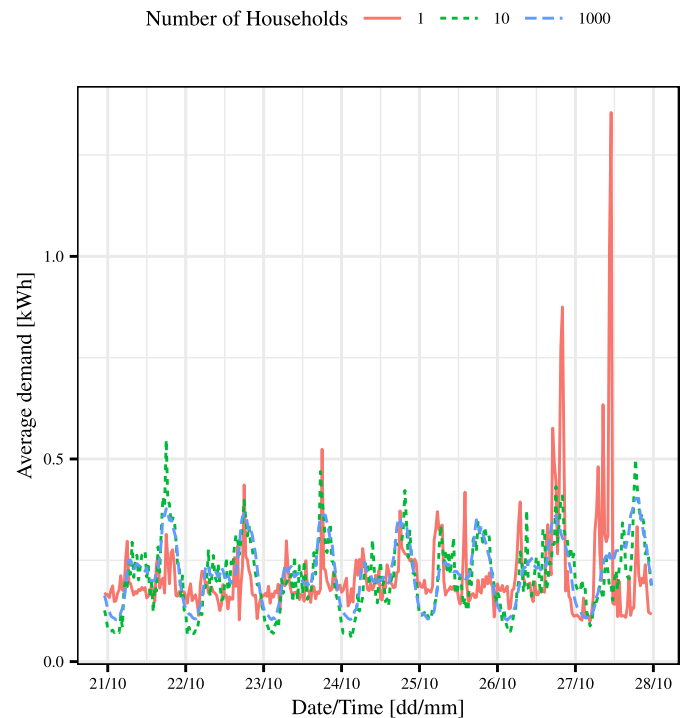


**Fig. 1.** Half-hourly demand during one week in 2013 averaged across an increasing number of households from 1 to 1000. Group behaviours become more apparent as the aggregation level increases.

network architecture [15]. In [16] several probabilistic methods (quantile regression, KDE) are evaluated using a dataset comprising of 100 real LV feeders where there was no clear best forecaster at all feeders, however, autoregressive type models performed well, and (forecast) temperature was shown to have negligible influence on the forecast skill.

Peak demand is typically the limiting factor in the capacity of distribution networks, set by the maximum power a cable or transformer can handle. Additionally, for a lot of flexibility applications the main goal is to reduce, flatten, or shift the daily peak demand in the LV network. Therefore, day-ahead forecasts of the daily maximum at the different nodes are valuable from both utilities perspective (e.g. in setting dynamic prices) and the consumers perspective (e.g. for scheduling battery or EV charging) [1]. The level of the daily peak is only half of the issue however, forecasting the time-of-peak is also relevant. There is little published in this specific area for the LV network to the best of our knowledge; related work [17] focuses on Extreme Value Theory and peaks over a defined threshold, which are by definition rare.

In this paper, we consider a four-level hierarchy: a primary substation (33 kV–11 kV), secondary substations (11 kV–4151 kV), feeders (415 kV), and households (230V, single phase). Methods for generating sharp and calibrated probabilistic forecasts of demand for the day-ahead are developed, including non-Gaussian parametric density forecasting at the household level, which is innovative in and of itself. We also investigate probabilistic daily peak forecasting, as in the daily maximum average energy demand, in terms of both peak intensity and timing. Finally, a method for combining (or blending) the daily bivariate (level and timing) peak forecasts and the half-hourly demand forecasts is described, which is termed forecast fusion. While methods for combining forecasts of varying spatial and/or temporal resolution have been proposed, combining forecasts of related yet distinct quantities has not been explored before, to the best of our

knowledge. The case study presented is based on a hypothetical network generated using the Low Carbon London dataset, and the proposed forecasts are robustly verified against benchmark models.

Probabilistic load forecasting is introduced in Section 2, followed by the concept of forecast fusion in Section 3. State-of-the-art methods for conventional day-ahead load forecasting are presented in Section 4.1, followed by proposed methods for daily peak intensity and timing forecasting in 4.2 and 4.3, respectively. An extensive, fully reproducible, case study based on the Low Carbon London dataset [18] and a hypothetical LV network is then presented where forecasts for load at household, feeder, secondary substation and primary substation levels are analysed in detail. Finally, brief conclusions are drawn in Section 6.

## 2. Probabilistic forecasting

In this section the probabilistic forecasting framework is formalised and the flexible statistical learning framework employed to generate forecasts throughout the work is introduced. We are typically interested in the predictive density or cumulative distribution function (CDF) of random variable $Y_t$ at time $t$. The predictive CDF is defined as

$$\hat{F}_t(y_t) = P(Y_t \leq y_t) \tag{1}$$

where $\hat{F}$ is a strictly increasing function. Here we consider time in half-hour periods, as this is the resolution of electricity metering in Great Britain, but other this is not a restriction on the methodology. A density forecast provides maximum flexibility as quantiles or intervals can easily be extracted from the forecast, and no need to approximation are necessary, such as interpolating between quantiles. Additionally, the full distribution may be described by a smaller number of parameters, which may be functions of explanatory variables. One drawback is that a suitable conditional parametric family must be found for the forecast. Kernel density estimation provides a non-parametric alternative but can be less flexible and more computationally demanding and therefore less scalable — critical for DSOs where the number of assets is large.

### 2.1. Generalised additive models for location, scale, and shape

Generalised Additive Models for Location, Scale, and Shape (GAMLSS) [19], are semi-parametric models. This is because a parametric distribution is assumed for the target variable, and the parameters that define the assumed distribution may depend on non-parametric smooth functions of explanatory variables. The framework is an extension of the more familiar Generalised Additive Model (GAM) [20], such that any parameter of the distribution can be a function of input features, not just the conditional mean.

If we have observations $y$, in this case demand at a particular location on the LV network, the conditional density typically $f(y|\boldsymbol{\theta})$ depends on up to four parameters; these are the location ($\theta_1$), scale ($\theta_2$), and shape parameters ($\theta_3$, $\theta_4$). An additive regression model is generated for each distribution parameter $\theta_i$ for $i = 1, \ldots, 4$. Let $\boldsymbol{x}_i$ be the pool of $N_i$ input features in the sub-model for $\theta_i$, and $g_i(\cdot)$ the link function, then the model formulation of a GAMLSS is

$$g_i(\theta_i) = \beta_{0,\theta_i} + \sum_{n=1}^{N_i} f_{n,\theta_i}(x_{i,n}), \quad i = 1, \ldots, 4 \tag{2}$$

where the function $f_{n,\theta_i}$ is the effect of explanatory variable $n$ on the distribution parameter $\theta_i$, which can be linear or non-linear functions, such as penalised smoothing splines, linear coefficients,
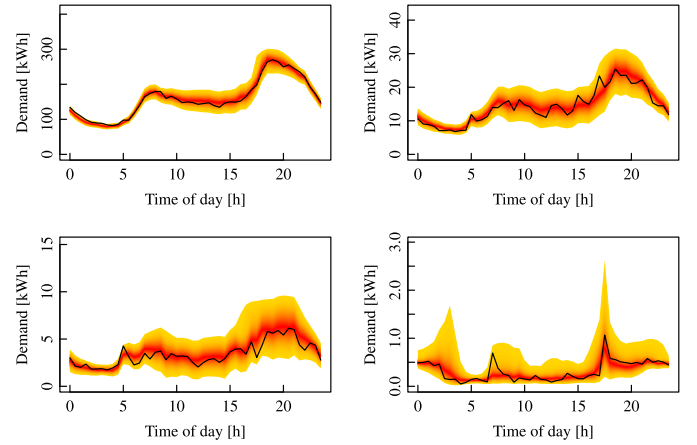


**Fig. 2.** Day-ahead example probabilistic forecasts as fan plots for the primary substation (`ps1` top left), a secondary substation (`ss1` top right), a feeder (`ss1_fdr1`, bottom left), and a household (`N1174`) on 17-10-2013. The widest and lightest coloured interval has a coverage probability of 98% and the measurement is overlaid in black for reference.

surfaces, etc; $\beta_{0,\theta_i}$ are the intercepts of each sub-model. These models may be estimated numerically using a combination of maximum likelihood, and successive back-fitting of the predictor functions for each parameter [19].

In Fig. 2 four example density forecasting models are visualised in a fan plot, where probability intervals are extracted from the conditional distribution at each lead time. Load forecasts are shown at different levels of aggregation. In general, forecast uncertainty is greater the further load is disaggregated. In particular, the possibility of large peaks is clearly quantified by high quantiles of the household-level forecast, which would not be captured by point forecasts.

Importantly, demand can approach zero in households although in our case study framework cannot be negative (net-demand, demand less embedded generation, is reserved for future work). This removes some parametric families for the forecasts from consideration, such as the Gaussian distribution. In Fig. 2, the Generalised Beta Prime distribution [21] is used for disaggregate demand and the Gaussian distribution is used for the three other aggregated levels. This small example is indicative of the approach throughout, which was to model the aggregate and household levels in the network distinctly; the forecasting models at the household level have to be simple, computationally efficient, and be suitable for right-skewed data.

## 3. Forecast fusion

In this paper, we hypothesise that better forecast skill can be achieved by fusing a bespoke forecast of the daily peaks in demand, in terms of both the timing and intensity of the event, with a state-of-the-art half-hourly resolution forecast, both produced one day-ahead. This technique is very similar to forecast combination (or blending, expert mixtures, etc.), but distinct in that we are combining forecasts of two different types: the demand for a particular half-hour and day $y_{d,h}$ and the daily peak in demand $y_d^{(p)}$. A forecast of the timing of the peak, i.e. the number of half-hours from midnight until the daily peak $h_d^{(p)}$, is also used in the method.

Consider the fusion of the two forecasts as a linear combination of the two distribution functions,

$$\tilde{F}_{d,h}(y) = (1 - w_{d,h})\hat{F}_{d,h}(y) + w_{d,h}\hat{F}_d^{(p)}(y) \tag{3}$$

with weights $w_{d,h}$ that may also be forecasts, or otherwise chosen, for each half-hour. This formulation respects the fact that a probabilistic forecast of the daily peak intensity (maximum value) is not sufficient in most applications and that the timing of the peak is also relevant. In this paper the weights are derived from probabilistic forecasts of the time-of-peak. If we define the random variable $H = \{1, 2, \ldots, 48\}$ to be the number of half-hours to each daily maximum in demand from midnight $h_d^{(p)}$, then the weights are found via a forecast probability mass function

$$w_{d,h} = \hat{f}_{d,h}(h) = P(H = h) \tag{4}$$

since forecasts are typically issued for discrete blocks of energy. Therefore, estimating the weights is re-framed into a discrete time-to-event prediction problem. Note that the sum of the probabilities $\hat{f}_{d,h}(h)$ over the total number of blocks in each day should be 1. If the conventional forecast were a prediction of load conditional on there not being a peak at $d, h$, this could be interpreted as a statement of the law of total probability. While it may be possible to produce such a forecast, we choose to proceed using conventional day-ahead forecasts so that the fusion method is as applicable as possible,

Forecast fusion is related to techniques found in probabilistic forecast combination [22,23], blending [24], expert mixtures [25], linear pools [26,27], and so on. Empirically it was found that important concepts in forecast combination, such as re-calibration of a linearly combined forecasts, were not necessary in the following case study. A beta-transformation of the combined forecast, following [26], reduced the forecast skill in cross-validation; further work on forecast combination may yield benefits here. However, all the aforementioned methods are based on combining forecasts of the same type (e.g. hourly resolution, day-ahead forecasts), hence the distinct terminology here, which highlights the similarity with the broader topic of data fusion.

Data fusion can be defined as the combination of multiple sources to obtain improved information in terms of quality, expense, and/or relevancy [28]. This is typically applied to combining data from multiple sensors and sources. In [29] multi-model forecast combination is discussed in terms of data fusion, however, the method entails combining forecasts of the same variable much like the literature discussed previously. In this work, we propose the fusion of conventional load forecasts with forecasts of peak load. These two forecasts exist in different temporal domains, hence this is a fusion problem not the usual practice of forecast combination or blending. We hypothesise that this will improve forecast skill overall as well as at peaks specifically, similar to how reconciliation of temporal hierarchies have been found to improve the skill across the different temporal domains [30].

## 4. Day-ahead load forecasting

In this section, the approach for generating the day-ahead probabilistic forecasts of half-hourly load, peak intensity and peak timing are detailed. Throughout, different approaches are used depending on the level of aggregation in the LV network, including model specifications and input features. The household level is treated as one group, and feeder level and above as the other, referred to as 'aggregated' levels. This prevents the framework from becoming too complex but is not a strict constraint. Common to all base probabilistic forecasts generated (the half-hourly forecasts $\hat{F}_{d,h}(y)$, the daily peak intensity forecasts $\hat{F}_d^{(p)}(y)$, and the peak timing forecasts $\hat{f}_{d,h}(h)$) is the GAM framework, where both half-hourly and peak demand forecasts utilise GAMLSS.

### 4.1. Half-hourly forecasts

The half-hourly forecasts at the aggregated levels of the LV network are described by $f_{d,h}(y|\theta_{1_{d,h}}; \theta_{2_{d,h}})$ where we assume the conditional distribution is Gaussian. The model is formulated as follows at each aggregate node

$$
\begin{aligned}
g_1(\theta_{1_{d,h}}) = &\beta_0 + \beta_1 y_{d-1,h} + \beta_2 y_{d-7,h} + \beta_3 y_{d-1}^{(p)} + \\
&\sum_{j=1}^{48} \alpha_j H_j(h) y_{d-1,h} + \sum_{j=1}^{48} \gamma_j H_j(h) y_{d-1}^{(p)} + \\
&f_{pvc}(h, D(d)) + f_{pbc}(d)
\end{aligned} \tag{5}
$$

and for the scale parameter, the formula is reduced to only depend on time-of-day for robustness

$$g_2(\theta_{2_{d,h}}) = \beta_0 + f_{pb}(h) \tag{6}$$

where $f_{pvc}(\cdot)$ is a varying coefficient penalised spline, $f_{pbc}(\cdot)$ is a cyclic penalised spline, and $f_{pb}(\cdot)$ is a penalised spline. There are two dummy variables for each half of the day $H_j(h)$, and period of the week $D(d)$ which is split into day type (Weekday, Saturday, and Sunday). So the load forecast is dependent on lags of the demand, a lag of peak demand, and interactions between yesterday's lags for each half-hour of the day. We also model the diurnal trend via the varying coefficient spline which changes according to day-type. Finally the annual seasonality is included, although in practice this spline is constrained to be very smooth to prevent interpolation of the data (in the case study we only have one year of data). Other formulations were tested; a full exploratory analysis can be found in the supplementary material [18]. However, this model formulation produced skilful forecasts relative to the benchmarks averaged over all time periods as well as during daily peaks, thanks to the interaction terms and the simple formula for the scale parameter.

Due to the complexity and sheer number of households, the half-hourly forecast models for the household level have to be more simple and robust than those at the aggregate levels. This is to prevent over-fitting and for computational efficiency. They are given by $f_{d,h}(y|\theta_{1_{d,h}}; \theta_{2_{d,h}}; \theta_{3_{d,h}}; \theta_{4_{d,h}})$ where we assume the conditional distribution follows the Generalised Beta Prime distribution. The model is formulated as

$$
\begin{aligned}
g_1(\theta_{1_{d,h}}) = &\beta_0 + \beta_1 y_{d-1,h} + \beta_2 y_{d-7,h} + \sum_{j=1}^{7} \alpha_j D_j(d) + \\
&f_{pb}(h) + f_{pbc}(d)
\end{aligned} \tag{7}
$$

and for the scale parameter, the formula is similar to the aggregated levels

$$g_2(\theta_{2_{d,h}}) = \beta_0 + f_{pb}(h) \tag{8}$$

and the two shape parameters of the distribution are constants to be estimated. Here the dummy variable $D_j(d)$ is for each of the 7 periods of the week.

### 4.2. Daily peak intensity

For the daily peak intensity forecasting, data exploration revealed seasonal trends and a high correlation in the lag dependency variables, as shown in Figs. 3 and 4 respectively, in the hypothetical LV network. Albeit the strength of the relationships are much weaker at the household level, again due to the low signal to noise ratio. An important consideration when creating the forecasting models here is the reduced size of the data set, since there is only one data point per day. Therefore, we reduce the number of features and categories in each formulation compared to half-hourly forecasting.
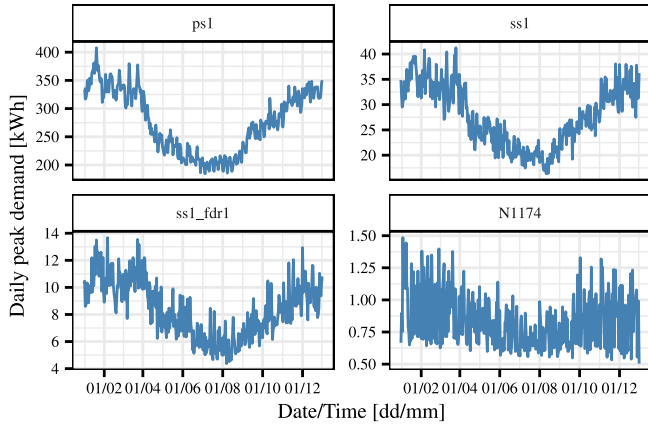
**Fig. 3.** Example time series of the daily peak intensity for primary substation (`ps1`), a secondary substation (`ss1`), a feeder (`ss1_fdr1`), and a household (`N1174`), from the hypothetical LV network. The peak demand shows seasonality at the aggregated levels, but again is more volatile at this household.
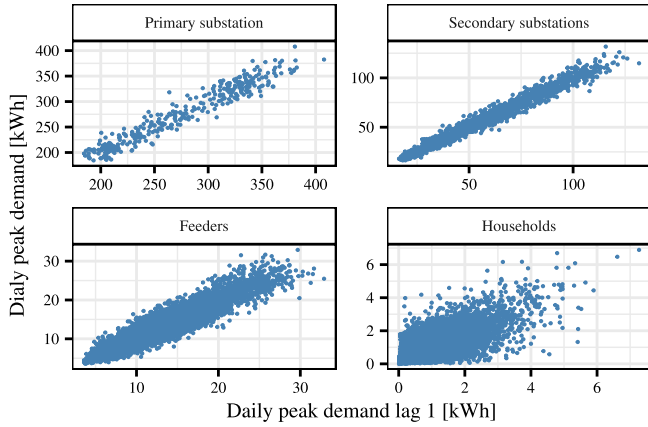


**Fig. 4.** Lag dependency plots of the daily peak intensity at four levels in a hypothetical LV network. This includes all nodes at each level and shows the motivation for using autoregressive based models for peak intensity forecasting, especially at the aggregate levels.

The peak intensity forecasts at the aggregated levels of the LV network are described by $f_d^{(p)}(y|\theta_{1_d}; \theta_{2_d})$ where we assume the conditional distribution is Gaussian. The model is formulated as follows at each aggregate node

$$g_1(\theta_{1_d}) = \beta_0 + \beta_1 y_{d-1}^{(p)} + \beta_2 y_{d-7}^{(p)} + \beta_3 \sigma_{y_{d-1}} + f_{pbc}(d) + \\ \beta_4 D_1(d) + \beta_5 D_2(d) \tag{9}$$

and for the scale parameter, the formula is

$$g_2(\theta_{2_d}) = \beta_0 + \beta_1 D_1(d) + \beta_2 D_2(d) + \beta_3 \sigma_{y_{d-1}} \tag{10}$$

where period of the week $D_i(d)$ is now reduced simply to either weekday or weekend categories, and $\sigma_{y_{d-1}}$ is yesterday's standard deviation of the half-hourly demand time-series.

An important feature at the household level for model robustness is the 'empty house' feature $I(d)$. We include lags of this variable, which defined an empty house by look at run-lengths of the daily standard deviation of the half-hourly demand. If this value dropped to approximately zero for a period of at least 7 days then the empty house feature is active. Note that the household had to have at least 30 days of emptiness for the feature to be included. This is because we want to identify houses here that are regularly empty for a reasonable period of time, rather than capture things such as holidays, public holidays, etc. which are
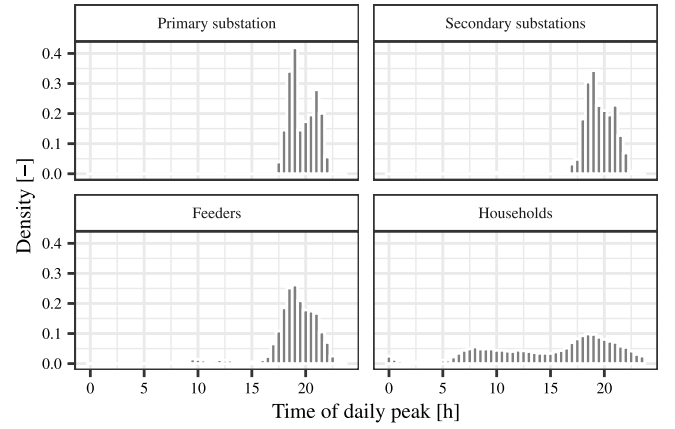


**Fig. 5.** Histograms of the daily peak timing at the four different aggregations of the hypothetical LV network. The peak timing becomes more dispersed at the lower aggregations on the network. At the primary substation level the peak timing during the analysis was consistently in the evening, except one data point which corresponds to the 'turkey peak' during Christmas day.

reserved for future work. So, the daily peak intensity forecasts at the household level are described by $f_d(y|\theta_{1_d}; \theta_{2_d}; \theta_{3_d}; \theta_{4_d})$ where we assume the conditional distribution follows the Generalised Beta Prime distribution. The model is formulated as follows at each household

$$g_1(\theta_{1_d}) = \beta_0 + \beta_1 y_{d-1}^{(p)} + \beta_2 y_{d-7}^{(p)} + f_{pb}(d) + \\ \beta_3 D_1(d) + \beta_4 D_2(d) + \beta_5 I(d-1) \tag{11}$$

and for $g_2(\theta_{2_d}) = \beta_0 + \beta_1 I(d-1)$, $g_3(\theta_{3_d}) = \beta_0 + \beta_1 I(d-1)$, the formula for the fourth moment of the distribution is kept constant. So the scale and shape parameters are only dependent on the empty house feature to make the forecasts more robust to overfitting.

### 4.3. Daily peak timing

A key component of the forecast fusion method is the weighting. In this study we choose to forecast the weights by defining them as the probability of the peak demand timing over the discrete blocks of energy in a day. Histograms showing the distribution of the peak timing at different levels on a hypothetical LV network are shown in Fig. 5, which shows that as expected the time of the daily peaks become more variable as demand becomes disaggregate. The timing of the peak demand, specially at the higher aggregations, is dependent on the time-of-year; it is widely understood during the winter the peak daily demand tends to be earlier in the evening (and the level of the peak becomes higher, an example of which is shown in Fig. 3) on the GB network. However, complex seasonal interactions were observed in the time-of-peak data, especially at the feeder and secondary substation levels of aggregation.

The process is framed as a discrete time-to-event problem, where with suitable transformations of the time-of-daily peak time series $h_d^{(p)}$, a GAM framework can be applied to generate forecasts [31]. This means we can leverage the powerful smoothing capabilities of a GAM to capture complex seasonal interactions between input features at the daily peak timing. The framework is relatively unique in terms of time-to-event or survival analysis, in that due to the framework there must be an event (i.e. peak) for every subject (i.e. day) for each experiment (i.e. node), and the domain of the peak timing in the analysis is $h_d^{(p)} \in \{1, 2, \ldots, 48\}$.
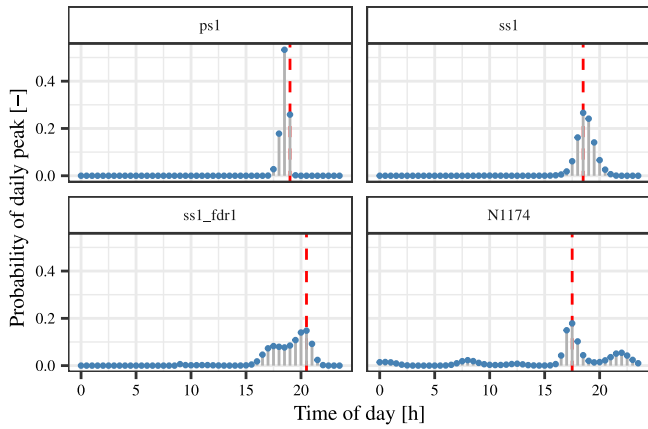
**Fig. 6.** Day-ahead example probability forecasts for the time of the daily peak at the same four nodes as Fig. 2 and for the same day. The measurement is represented by the dashed red vertical line in each panel for reference. These forecasts correspond to the weights used in the forecast fusion for each discrete time period of this day.
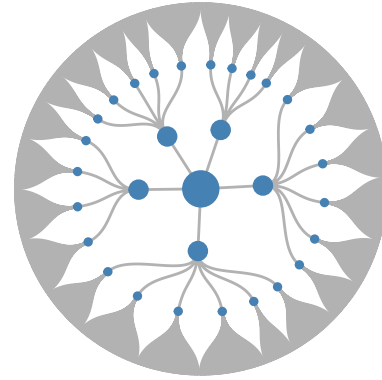


**Fig. 7.** Diagram of the hypothetical LV network hierarchy used in the case study where the central node represents the primary substation and so on down to the household level at the end nodes. The size of the primary substation is small compared to reality in the GB distribution network due to the limited number of smart meters used. However, as shown in Fig. 1, group behaviours begin to emerge quickly at high levels of aggregation.

The discrete hazard function is the conditional probability at time interval $h$

$$\lambda(h|x) = P(H = h|H \geq h, x) \qquad (12)$$

which gives the conditional probability of a peak in interval $h$ given that the peak happens at time $H \geq h$. To model this hazard rate, we use a time varying linear predictor with a GAM framework

$$\lambda(h|x) = g^{-1}\left[\beta_{0h} + \sum_{n=1}^{N} f_n(x_{nh})\right] \qquad (13)$$

where the link function here $g(\cdot)$ is the logit link. In this case for the aggregated levels in this case we use a tensor interaction term for the period of day and day of year as the input features as well as a dummy variable for the day type (weekday/weekend). At the household level a more simple approach is needed for computational efficiency; two separate smooth splines for each of the input features are used. The discrete survival function is defined as

$$S(h|x) = P(H > h|x) = \prod_{s=1}^{h}(1 - \lambda(h|x)) \qquad (14)$$

from which the discrete cumulative distribution function is found $F(h) = 1 - S(h)$, and then the probability mass function for each day $f_h(h) = P(H = h)$ is easily calculated. For estimating the regression model, an indicator variable for each daily recorded value of $h^{(p)}$ is defined as the target variable which is zero for each discrete block of time until the event is recorded where the indicator variable is 1. For more information on the data transformation the reader is referred to [32,33]. Note that the final peak timing probability forecast could similarly be obtained via a simple logistic regression/classification method, but the discrete time-to-event setting gives the proposed method a stronger theoretical foundation. In Fig. 6 four probability forecasts are shown for the same LV nodes as Fig. 2 and for the same day; this constitutes the weights $w_{d,h}$, i.e. the last component required to fuse the half-hourly and daily peak intensity forecasts together.

## 5. Case study

The methodology is tested using data from the Low Carbon London trial [34]. Households which are on a variable price tariff

are first removed, and then households which have regular communication issues and/or suspect data are removed, and finally only households which have a complete record of measurements during 2013 are retained. This represents quite a strict data cleaning process in which we are left with 742 smart meters; however, addressing challenges due to missing-data and forecasting dynamic price tariff households are beyond the scope of this study. Data from this experiment is anonymised which means location specific effects, such as temperature, are necessarily excluded from the analysis.

The households are sampled (without replacement) to create a hypothetical LV network, albeit given the amount of the available smart meters, the primary substation level (i.e. the top aggregation) is smaller than perhaps you would find in practice. However, as shown in Fig. 1, group behaviours begin to emerge quickly. The sampling process was configured whereby each secondary substation comprises of between 4–7 feeders and each feeder contains around 16–45 smart meters. The hypothetical LV network is illustrated in Fig. 7 showing all the different levels of aggregation. As discussed in [4] there is a severe lack of real open-access LV network data available. Therefore, this approach is effectively a compromise because we implicitly cannot account for street furniture, embedded generation, and potential correlations between nodes in a real life network.

For the regression problems, the coverage of the dataset is January 2013 to December 2013 inclusive. To compare the forecast performance, each month is partitioned into thee blocks of approximately 10 days, depending on month. The first two blocks of every month compromise the training data, on which cross-validation is also carried out to generate out of sample forecasts covering the full year. The last block in each month constitutes the testing data, i.e. not used anywhere for model estimation or tuning. All methods are implemented in R [35] using the package ProbCast [36], developed by the authors for the modelling, evaluation, and visualisation of probabilistic forecasts. The wrapper functions for GAMLSS for the regression models [19] are used extensively here.

A two pronged approach is necessary to assess probabilistic forecasts performance; calibration (or reliability) is the necessary condition that predicted probabilities are unbiased, and sharpness is a measure of forecast uncertainty, i.e. the spread of the predictive distribution. Combined, these allow for the ranking of competing forecasting methods by sharpness subject to calibration [37]. For the full predictive cumulative distributions in this

case study, the Continuous Ranked Probability Score (CRPS) [37, 38] is used to measure both sharpness and calibration

$$\text{crps} = \frac{1}{N} \sum_{t=1}^{N} \int_{-\infty}^{\infty} \{\hat{F}_t(y) - \mathbf{1}(y \geq y_t)\}^2 dy \tag{15}$$

where $\mathbf{1}(\cdot)$ is the indicator function. The CRPS for a single forecast observation pair is therefore the area between the squared difference of the forecast and observation CDF, where the latter is a step-function from 0 to 1 at the observed value. For the discrete probability forecasts of the peak timing, the Ranked Probability Score (RPS) is used for verification, which is the discrete form of CRPS [39].

The Probability Integral Transform (PIT) histogram is used to verify the calibration of the forecasts

$$u_t = \hat{F}(y_t) \tag{16}$$

whereby, if the forecasts are well calibrated and the sample is sufficiently large, $u \sim \mathcal{U}(0, 1)$, which is inspected visually via a histogram with a certain number of (typically 20 [39]) bins. One limitation with this approach is that is becomes time consuming to verify all nodes in a large network. So, in this case the calibration checked within grouped network levels (e.g. smart meters, feeders, etc.). Reliability diagrams [40] are used as an alternative to check the calibration of quantiles of the distribution at individual nodes in the supplementary material [18]. The calibration of the discrete probability forecasts is not presented here for brevity, but are shown in the supplementary material via multi-category reliability diagrams [18,41].

### 5.1. Benchmarks & skill scores

A set of benchmark models have been chosen for case study, with models chosen depending on the forecasting task and level in the LV network. There a limited literature on benchmarks for probabilistic forecasts in general, and none established for LV load forecasting.

For the half-hourly forecasts a simple auto-regressive model is used and implemented in the GAMLSS framework for the aggregated levels

$$g_1(\theta_{1d,h}) = \beta_0 + \beta_1 y_{d-1,h} + \beta_2 y_{d-7,h} + f_{pb}(h) \tag{17}$$

with and the scale parameter and distribution family the matching the proposed advanced model. At the household level three benchmark models are employed; two are based on Kernel Density Estimation (KDE) using zero-truncated Gaussian kernels. In the first separate KDE models are defined for each for each half-hour of the day, and in the second separate KDE models are defined for each half-hour and day type (weekday, Saturday, Sunday). Both of these are similar to a previously published method [9]. The third is a simplified version of the GAMLSS model used at the household level, with the same Generalised Beta Prime family, where only the location and scale parameters vary smoothly with time-of-day.

For the peak intensity forecasts a very simple autoregressive based GAMLSS model is used for the aggregated levels, with the Gaussian conditional distribution and formula

$$g_1(\theta_{1d}) = \beta_0 + \beta_1 y_{d-1}^{(p)} + \beta_2 y_{d-7}^{(p)} \tag{18}$$

where the scale parameter is a constant. At the household level, again three benchmark models are employed. The first is a very simple unconditional KDE estimate, the second is a KDE estimate conditional on day type (weekday/weekend), and finally a simple location-only autoregressive GAMLSS model, i.e. the same as Eq. (18) except using the Generalised Beta Prime family as the conditional distribution.
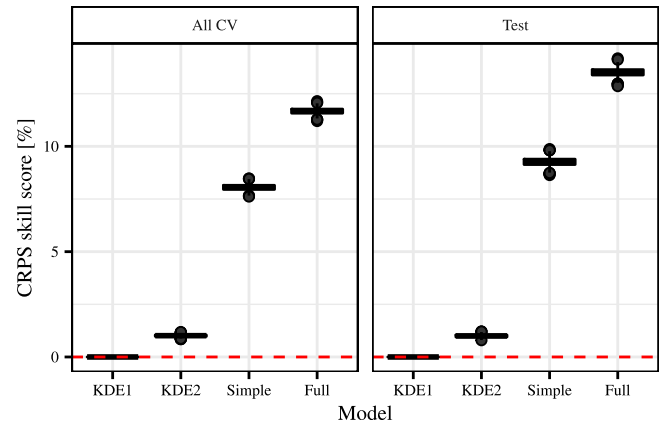


**Fig. 8.** Skill scores averages of the peak intensity forecasts at the household level of the network relative to KDE1, where the Full model is a component of the Fusion model. The sample distribution is found via bootstrap averages. The peak intensity forecasts at the household level are as skilful as those at the aggregate levels relative to the benchmark.

For the peak timing probability forecasts the same simple seasonal climatology model is used at all points of the network. This is a competitive benchmark because seasonality is the only effect used in our more advanced model.

Forecast evaluation is reported via the relative change of the score of the proposed model $\bar{S}$ to a benchmark $\bar{S}^{ref}$ via skill scores. If the perfect score is zero, as in the cases considered here, then the skill score is

$$\text{skill} = 1 - \frac{\bar{S}}{\bar{S}^{ref}} \tag{19}$$

and in the following the terms skill score, percentage improvement, and relative change are used interchangeably. Bootstrap re-sampling is used as a simple non-parametric method for estimating the significance in forecast improvement [40]. Forecast target times are sampled with replacement with a length equal to the original length of the set, and then skill scores are calculated. This process is repeated a large number of times until the sampling variation of the result is determined, which are then typically presented via boxplots [40].

### 5.2. Daily peak intensity evaluation

In Fig. 8 the skill of the peak intensity forecasts is demonstrated at the household level against three benchmarks, an unconditional KDE, KDE conditional on day type (weekday/weekend), and a simple GAMLSS model, based on autoregressive features for the location parameter only. The full model is used as a component in the fused forecasts, and as you can see on average results in improved forecasts of over 15% relative to the most simple benchmark in testing. This validates the motivation for using a bespoke model for predicting the peaks alone. There is some variability in the skill between households and this is skewed towards improved skill, as detailed in [18] and discussed in Section 5.5. Some households show over a 50% improvement in CRPS in cross-validation and testing compared to the unconditional KDE estimate and few showing negative skill below −5%. The improvement between the two GAMLSS models validates the inclusion of the smooth day of year, day type (weekday/weekend), and empty house features.

The skill scores are similar across all levels of the network from Feeder to Primary Substation, in contrast to peak timing and half-hourly forecasts which are generally more skilful at higher levels of aggregation than lower. The breakdown by individual nodes is

presented in [18] and we find that skill is positive and statistically significant in the majority of cases on the test data, but also note that the sample variation is greater due to the reduced data volume. This is also reflected in how the nodes where Fusion shows the greatest/least skill are not the same between the cross-validation and test results, although the overall behaviour is similar.

### 5.3. Daily peak timing evaluation

In this subsection we evaluate in more detail the weights used in the fusion forecast, $w_{d,h}$, which is defined as the probability of the daily peak at each node occurring in each discrete block of energy throughout the day. At the aggregate levels, the skill scores of the discrete time-to-event probability forecasts, where the reference is seasonal climatology, are range from 0% (for four feeders) to over 20% for the Primary Substation. Detailed results may be found in [18]. There is a clear trend for greater improvement at higher levels in the hierarchy. This is because the peak timing is less variable and more smoothly dependent on seasonal, which is easily modelled. Whereas, even at the feeder level during testing some of the nodes have similar or less skill than the benchmark, although there is a skew generally for positive skill scores. At the feeder levels the time of peak is clearly dependent on more complex behaviours than can be described by time of day and day of year. Additionally, the general change in skill between cross-validation and testing is worth further investigation at all the levels; ideally more data would be available when learning complex seasonal interactions between day of year and time of day which may be leading to overfitting.

This problem is also evident at the smart meter level. The time-to-event based forecast is only marginally more skilful in testing than the benchmark with a CRPS skill score of under 0.5%. This highlights the difficulty in predicting the peak timing at the household level and that the predictions of the time-to-event based model are close to seasonal climatology on average. The variability in skill has been investigated, shown in [18]. There is variation in skill between households, which ranges from ±10%; some households are apparently more predictable than others.

Teasing out relationships between node type, the skill of the timing probability forecasts, and drivers of predictability is an interesting aspect of future work. Stakeholders could use this information to gauge locations for the provision of flexibility in the LV network. Finally, further analysis of the forecasts could give an understanding as to which nodes are likely to reach their daily peak at the same time as the more aggregated nodes higher up the network. This information could be valuable for revealing which households or nodes to leverage for peak demand shifting via (for example) time-of-use tariffs.

### 5.4. Forecast fusion evaluation

In the following subsections we first evaluate the forecast fusion methodology. To this end, we evaluate the forecasts using CRPS skill scores averaged over all time periods, and also inspect the calibration directly. However, to demonstrate the improved forecast skill for the daily peak demand we also retrospectively select periods where the daily peak demand is recorded and evaluate the fused forecast averaged over these time periods only. Due to the double penalty effect [7], an advanced (and smoother) forecast might produce improved skill on average compared to a benchmark, but fail predict peak demand well. Therefore, the structure of this evaluation is aimed to demonstrated the skill of the forecast on average and during the daily peak demand.
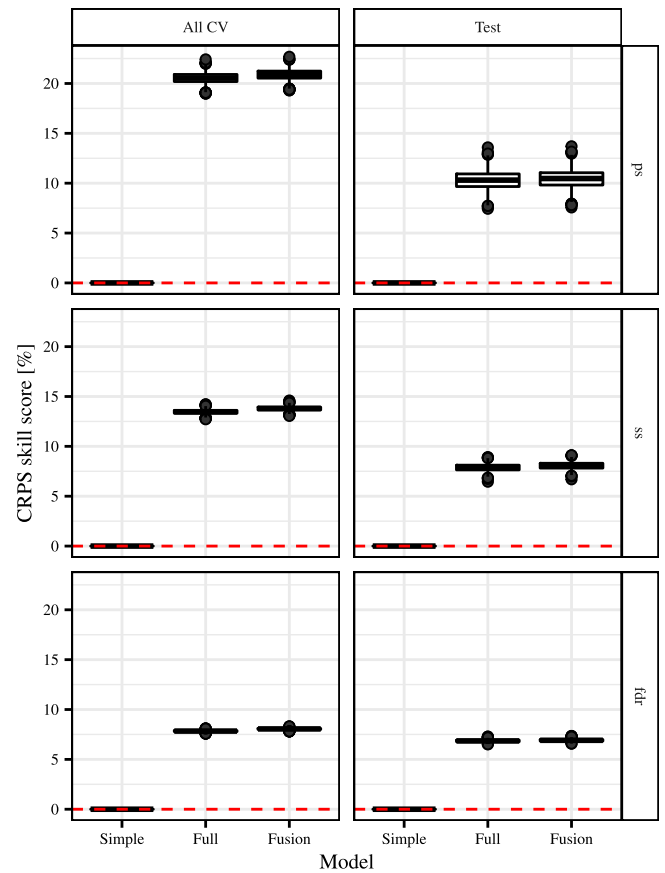


**Fig. 9.** Skill scores averages of the three half-hourly forecasting methods employed at the aggregated levels of the network relative to the 'Simple' model. The sample distribution is found via bootstrap averages, where all available samples are included at the primary substation (ps), secondary substation (ss), and feeder (fdr) levels.

### 5.4.1. Aggregate levels

At the aggregate levels, the two advanced forecasting methods show similar improved skill over the benchmark at the primary substation (ps), secondary substation (ss), and feeder (fdr) levels in the network. This is true for both cross-validation and testing, as shown in Fig. 9. So we can conclude that the forecast fusion method is at least as skilful as the full half-hourly model. This is encouraging since the latter is a key component in the fused forecast. More generally, we can see the skill improvements possible from adding seasonal features and interactions in the regression model, as evidenced by the ≈10% improvement in forecast skill during testing across the aggregated network compared to the simple autoregressive benchmark.

Rather than metrics at each level of the network, in Fig. 10 the skill scores of the Fusion method are plotted at each single node of the aggregated network. Clearly there is variability in improvement between nodes, especially at the feeder level which is hidden in Fig. 9. This plot also emphasises that during cross-validation the improvement is increasing with the aggregation (or voltage) level. However, during testing the skill scores are far closer between the levels which could indicate that the testing data is more difficult to predict at the aggregate levels, the benchmark models are improved relative to the advanced models with more training data, or the advanced models are over-fitting slightly, or a combination of all three.

If we evaluate the forecasts only during the periods when the daily peak demand is recorded the skill of the three forecasting methods looks very different. Fig. 11 shows that the full model
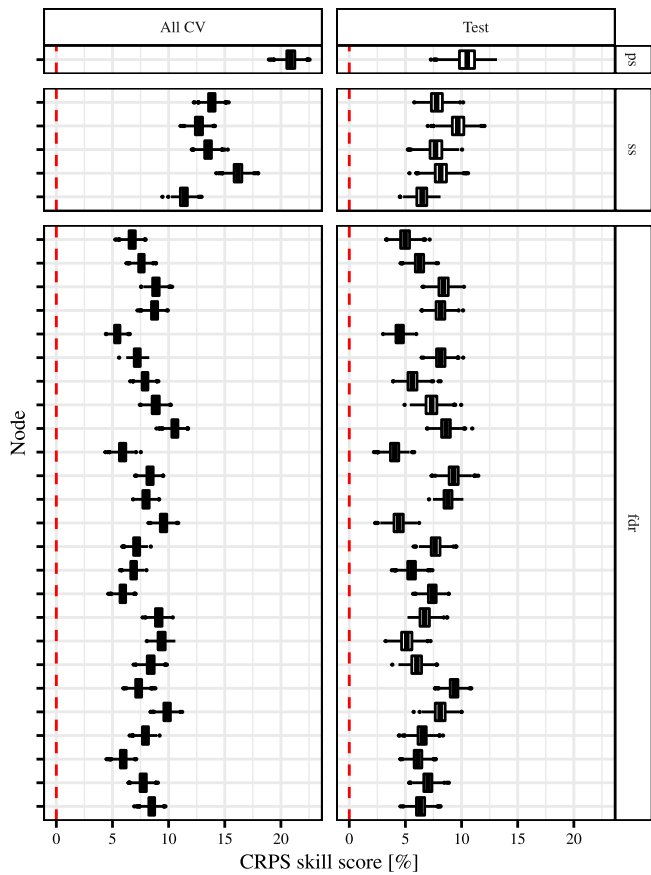
**Fig. 10.** Skill scores averages of the Fusion forecasting method at the aggregated levels of the network relative to the 'Simple' model. The sample distribution is found via bootstrap averages, where all available samples are included at each node.



**Fig. 11.** Skill scores averages of the three half-hourly forecasting methods employed at the aggregated levels of the network relative to the 'Simple' model. The sample distribution is found via bootstrap averages, where only samples which correspond to the daily peak demand are included at the primary substation (ps), secondary substation (ss), and feeder (fdr) levels.



**Fig. 12.** PIT histograms of two forecast models at the aggregated levels of the network. Note that except from the primary substation (ps) level, these histograms show the average calibration of all the nodes. The fused forecast is similarly calibrated at all levels and across both data partitions.

is similar in skill during testing to the benchmark. In fact, during the data exploration and tuning of the models, it proved difficult to find a suitable feature set which performed equally well to the benchmark model during the peak half-hours, as evidenced in the supplementary material [18]. However, the Fusion method is again ≈10% better at forecasting the daily peak during testing across the aggregated network compared to the benchmark. This is a key result of the paper and shows that by fusing a bespoke forecast of the daily peaks to a state-of-the-art half-hourly forecasts, it is possible to achieve skilful forecasts on average *and* during the daily peak demand. We have also investigated the variation in forecast performance across individual nodes of the aggregated network. While there is again variability between nodes, all show improvement from forecast fusion of between 5% and 15%. Bootstrap skill scores by node are illustrated in [18].

The calibration of the advanced GAMLSS and fusion models is shown in Fig. 12 where essentially the average calibration at the aggregated levels is shown. Importantly, the Fusion method is at least as well calibrated as the Full model, if not marginally better calibrated. Clearly, the right tail of the distribution could be improved at most of the levels to account for large peaks in demand. However, this is reserved for future work. Additionally, the calibration of both models at the feeder level is relatively poor on average. Using distribution free regression approaches might be beneficial here as well as accounting for holiday and special events. The key result here is that the forecasts are reasonably well calibrated and the fusion methodology did not introduce calibration issues via the linear combination, an issue that is widely discussed in forecast combination [26].
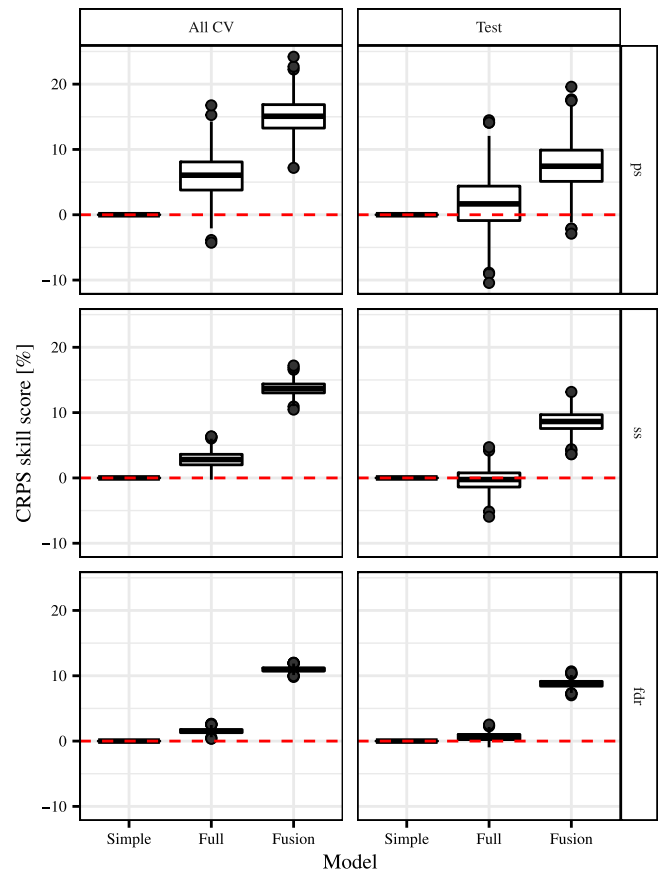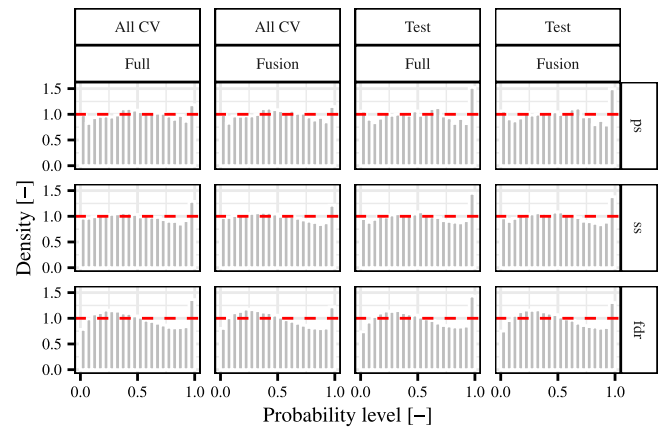
### 5.4.2. Household level

Recall that there are three benchmarks for household level forecasts, two variations on KDE and a very simple GAMLSS model based only on time-of-day. An interesting result, shown in Fig. 13 is that using autoregressive and seasonal terms it is possible to achieve better skill than simple benchmarks, by 3%–4% during testing in this case study. As well, the Fusion
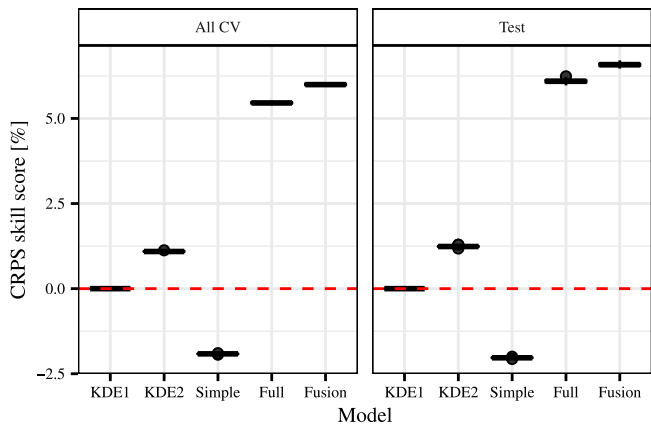
**Fig. 13.** Skill score averages of the five half-hourly forecasting methods employed at the household level of the network relative to KDE1. The sample distribution is found via bootstrap averages, where all available samples are included.



**Fig. 14.** Skill scores averages of the Fusion forecasting method at the household level of the network during peaks relative to KDE1. The sample distribution is found via bootstrap averages, using only samples which correspond to the daily peak demand are included at each node.

methodology is marginally better than the advanced GAMLSS model in both cross-validation and testing. The Simple GAMLSS model is the worst performing out of the methods tested. At some nodes (8 out of 742) the Full model failed to converge and the performance of the resulting forecasts was very poor during cross-validation. Further inspection revealed issues at these households such as structural changes in the time-series and so on. At these 8 nodes, the Simple model was therefore used in place of the Full. Additional detail on this can be found in the supplementary material [18].

As one would expect, the relative performance of the benchmark and Fusion methods varies between households. Unlike the aggregated level where there was consistent improvement at all nodes for the proposed Fusion methodology, at the household level forecast fusion provides improvement for 80% of households relative to KDE1, and 70% relative to the sophisticated Full GAMLSS benchmark. This is due to the diversity of behaviours at the household level and perhaps at some nodes the model is imply over-parameterised given the information in the time series which suggests perhaps boosting or regularisation would be beneficial in the model fitting if computationally feasible. However, the density is clearly skewed towards improvement as you would expect from Fig. 13. Supplementary results on this topic are provided in [18].

Again, when evaluated during the periods when the daily peak demand is recorded the ranking of the different forecasters is very different. As shown in Fig. 14, the simple and advanced GAMLSS models are now significantly worse than even the simple time-of-day KDE benchmark by ≈1%–5% during testing, which demonstrates the double penalisation effect reported in the literature. However, the fusion model remains the most skilful forecast, and is significantly better than the more advanced KDE model averaged over both cross-validation and testing. Although the skill score is not as large as the aggregated levels detailed above, there is still a large and significant improvement from the full model, which is one of the inputs to the fused forecast, of approximately 6% during testing.

In terms of calibration, the PIT histogram for all the households is shown in Fig. 15. Although it is not possible to distinguish individual nodes in this case the plot shows that both the advanced GAMLSS and forecast fusion method are reasonably well calibrated across this level of the hierarchy, with some evidence of over-confidence. However, given the nature of smart meter demand and that we are using a parametric assumption
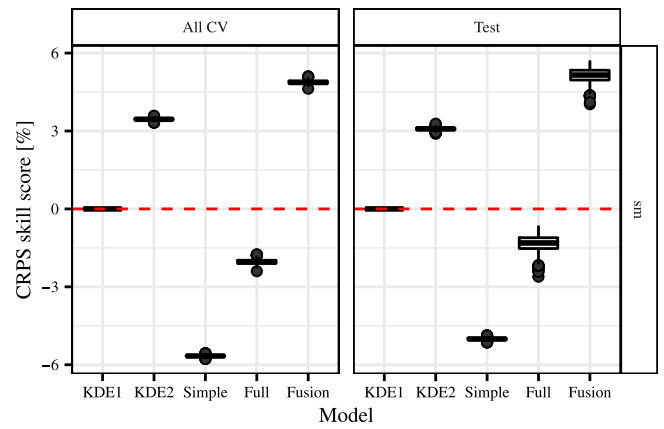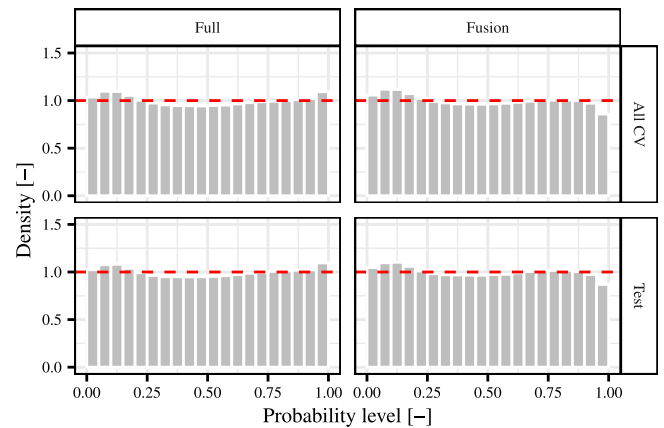


**Fig. 15.** PIT histograms of two forecast models at the household level of the network, which show the average calibration of all the nodes. The fused forecast is similarly calibrated at all levels and across both data partitions, except for the right tail of the two methods.

for the predictive distribution, the calibration is better than expected. There are some differences between the calibration of the GAMLSS and Fusion forecast now however, with the right tail of the distribution going from too narrow to too wide on average. This indicates a possible area of improvement for the forecasts.

### 5.5. Results summary and discussion

To summarise, at aggregated and household levels, fusion of peak and conventional forecasts provides a significant improvement in CRPS relative to both simple and advanced benchmarks methods during peak hours. The improvement relative to advanced methods ranges from 6% to 9%. This comes with no penalty to overall performance, which is also improved but only marginally. Average CRPS is reported for each aggregation level in Table 1, though we refer the interested reader to the earlier figures and supplementary material where we verify that this improvement is largely consistent across individual households, feeders and substations, which differ in size and variability.

Finally, we have investigated whether there is any relationship between the variability of load (at substations, feeders and households) and forecast improvement of the fusion method relative to benchmarks. We have compared the skill scores with the coefficient of variation for all individual substations, feeders

**Table 1**

Summary of the mean CRPS from the test set for the Simple and Advanced (Adv.) half-hourly forecasts, and Fusion forecasts. Scores are presented for All periods and Peak periods. Skill scores are for the Fusion forecasts relative to the Advanced benchmark.

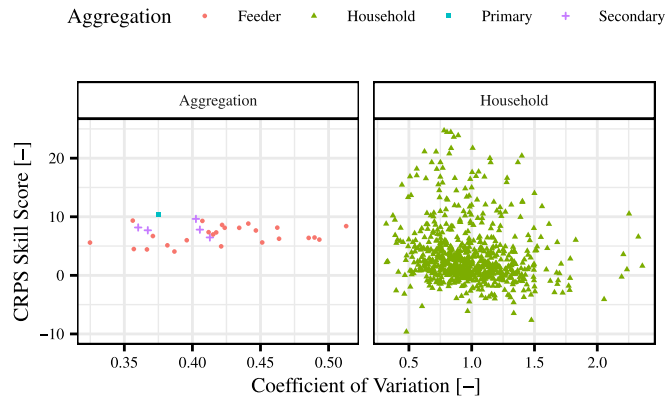|              | Aggregation | Simple | Adv. | Fusion | Skill |
|--------------|-------------|--------|------|--------|-------|
| All time     | Primary     | 7.04   | 6.32 | 6.31   | 0.2%  |
|              | Secondary   | 1.92   | 1.77 | 1.77   | 0.2%  |
|              | Feeder      | 0.69   | 0.64 | 0.64   | 0.0%  |
|              | Household   | 0.08   | 0.08 | 0.08   | 0.4%  |
| Peaks only   | Primary     | 8.94   | 8.80 | 8.30   | 5.7%  |
|              | Secondary   | 3.08   | 3.09 | 2.81   | 9.0%  |
|              | Feeder      | 1.54   | 1.53 | 1.40   | 8.2%  |
|              | Household   | 0.47   | 0.47 | 0.44   | 6.0%  |



**Fig. 16.** Forecast skill of the fusion forecast relative to benchmark (KDE1 for households, Simple half-hourly forecast for aggregations) against coefficient of variation for each substation, feeder and household.

and households, illustrated in Fig. 16. There is no apparent relationship between variability and skill for any aggregate level, which all have positive skill. Of course we have few examples of primary and secondary substations, but skill at these levels is comparable to individual feeders. For households, however, we observe a negative correlation between variability and forecast skill, although there is a large amount of variation, and positive skill for 80% of households. Furthermore, only households with a relatively low coefficient of variation exhibit very high forecast skill. This highlights the importance of considering forecast skill for individual households, as 'average' performance across multiple households will mask this variation in forecast performance. For most applications, we believe that significant improvement at 80% of households is more than sufficient to justify a slight increase in the complexity of the forecasting process.

## 6. Conclusions

Forecasting methods that are effective across all voltage levels of distribution networks will be essential as Distribution Network Operators take on new responsibilities for managing energy balancing and ancillary services. This paper presents a novel approach to probabilistic load forecasting that addresses deficiencies of existing methods caused by peak loads, and is shown to improve forecast skill across distribution networks from household to primary substation level by as much as 10% overall, and more during peaks. The skill of forecasts during peaks is particularly important in many use-cases, including network constraint management, peak shaving, and battery and demand response scheduling.

The approach we propose combines forecasts of daily peak timing and intensity with conventional load forecasts. By forecasting peaks specifically, we compensate for the tendency of

conventional methods to be too smooth and under-forecast peaks. Probabilistic forecasts are combined or 'fused' using a simple weighting scheme inspired by the more general practice of data fusion. A comprehensive case study based on open data is presented where we find that while sophisticated methods for conventional forecasts may provide skill overall with respect to competitive benchmarks, they add little value during peaks. Fusion of conventional forecast with a peak forecast marginally improves performance overall, and greatly improves performance during peaks. Average improvement during peaks ranged from 6% at household and Primary substation level, and 8%–9% at Feeder and Secondary substation level in our case study.

Additionally, we have proposed a method for producing parametric density forecasts at the household level based on the Generalised Beta-Prime distribution and the GAMLSS framework. This is in contrast to non-parametric methods that have dominated the literature to date and are far less parsimonious. The proposed method produced an average improvement of approximately 5% relative to methods based on Kernel Density Estimation across the 742 households in our case study.

However, forecasting capabilities require further development to meet the expected future needs of DSOs. Not least, consideration should be given to embedded generation, storage, and demand response. Furthermore, to be of maximum practical use, forecasting models should be applicable to feeders/substations they have not been trained on, known as 'global' forecasting models as proposed in [42]. As distribution networks feature tens-of-thousands of LV feeders, the use of domain adaptation via transfer learning would take a pre-trained model and adapt it to any feeder given minimal adjustment, but certainly not retraining. Development of forecasting models that are adaptive to track changes in load behaviours, structural breaks in particular, should also be considered. Another aspect to consider and potentially exploit is the hierarchical nature of electricity demand. Encoding this structure in forecasting models can help improve accuracy and enable more coordinated decisions at different levels of the network.

**CRediT authorship contribution statement**

**Ciaran Gilbert:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Jethro Browell:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Visualization, Writing – review & editing. **Bruce Stephen:** Conceptualization, Data curation, Funding acquisition, Project administration, Supervision, Writing – review & editing.

**Declaration of competing interest**

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jethro Browell reports financial support was provided by Engineering and Physical Sciences Research Council. Bruce Stephen reports financial support was provided by Engineering and Physical Sciences Research Council. Jethro Browell is a guest editor for the SEGAN special issue on "Forecast production and end-use for efficient management of energy systems".

**Data availability**

All data is available in a linked Zenodo repository.

## Acknowledgements

## References

[1] G. Mokryani, Transition from distribution network operator to distribution system operator, Future Distrib. Netw. (2022) 1–16, http://dx.doi.org/10.1063/9780735422339_010.

[2] T. Hong, S. Fan, Probabilistic electric load forecasting: A tutorial review, Int. J. Forecast. 32 (3) (2016) 914–938.

[3] J. Zhu, H. Dong, W. Zheng, S. Li, Y. Huang, L. Xi, Review and prospect of data-driven techniques for load forecasting in integrated energy systems, Appl. Energy 321 (2022) 119269, http://dx.doi.org/10.1016/J.APENERGY.2022.119269.

[4] S. Haben, S. Arora, G. Giasemidis, M. Voss, D. Vukadinović Greetham, Review of low voltage load forecasting: Methods, applications, and recommendations, Appl. Energy 304 (2021) http://dx.doi.org/10.1016/j.apenergy.2021.117798.

[5] B. Yildiz, J.I. Bilbao, J. Dore, A.B. Sproul, Recent advances in the analysis of residential electricity consumption and applications of smart meter data, in: Applied Energy, vol. 208, 2017, http://dx.doi.org/10.1016/j.apenergy.2017.10.014.

[6] Y. Wang, Q. Chen, T. Hong, C. Kang, Review of smart meter data analytics: Applications, methodologies, and challenges, IEEE Trans. Smart Grid 10 (3) (2019) http://dx.doi.org/10.1109/TSG.2018.2818167.

[7] S. Haben, J. Ward, D. Vukadinovic Greetham, C. Singleton, P. Grindrod, A new error measure for forecasts of household-level, high resolution electrical energy consumption, Int. J. Forecast. 30 (2) (2014) http://dx.doi.org/10.1016/j.ijforecast.2013.08.002.

[8] R.J. Bessa, C. Möhrlen, V. Fundel, M. Siefert, J. Browell, S. Haglund El Gaidi, B.-M. Hodge, U. Cali, G. Kariniotakis, Towards improved understanding of the applicability of uncertainty forecasts in the electric power industry, Energies 10 (9) (2017) http://dx.doi.org/10.3390/en10091402.

[9] S. Arora, J.W. Taylor, Forecasting electricity smart meter data using conditional kernel density estimation, Omega (United Kingdom) 59 (2016) http://dx.doi.org/10.1016/j.omega.2014.08.008.

[10] M. Reis, A. Garcia, R.J. Bessa, A scalable load forecasting system for low voltage grids, in: 2017 IEEE Manchester PowerTech, IEEE, 2017, pp. 1–6, http://dx.doi.org/10.1109/PTC.2017.7980936.

[11] Y. Wang, D. Gan, M. Sun, N. Zhang, Z. Lu, C. Kang, Probabilistic individual load forecasting using pinball loss guided LSTM, Appl. Energy 235 (2019) http://dx.doi.org/10.1016/j.apenergy.2018.10.078.

[12] S. Ben Taieb, R. Huser, R.J. Hyndman, M.G. Genton, Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression, IEEE Trans. Smart Grid 7 (5) (2016) http://dx.doi.org/10.1109/TSG.2016.2527820.

[13] M. Arpogaus, M. Voss, B. Sick, M. Nigge-Uricher, O. Dürr, Short-term density forecasting of low-voltage load using Bernstein-polynomial normalizing flows, 2022, http://dx.doi.org/10.48550/arxiv.2204.13939, arXiv.

[14] S.B. Taieb, J.W. Taylor, R.J. Hyndman, Hierarchical probabilistic forecasting of electricity demand with smart meter data, J. Amer. Statist. Assoc. (2020) 1–17, http://dx.doi.org/10.1080/01621459.2020.1736081.

[15] R. Telford, B. Stephen, J. Browell, S. Haben, Dirichlet sampled capacity and loss estimation for LV distribution networks with partial observability, IEEE Trans. Power Deliv. 36 (5) (2021) http://dx.doi.org/10.1109/TPWRD.2020.3025125.

[16] S. Haben, G. Giasemidis, F. Ziel, S. Arora, Short term load forecasting and the effect of temperature at the low voltage level, Int. J. Forecast. 35 (4) (2019) http://dx.doi.org/10.1016/j.ijforecast.2018.10.007.

[17] M. Jacob, C. Neves, D. Vukadinović Greetham, Forecasting and assessing risk of individual electricity peaks, in: Mathematics of Planet Earth, Springer International Publishing, Cham, 2020, pp. 39–60, http://dx.doi.org/10.1007/978-3-030-28669-9.

[18] C. Gilbert, J. Browell, B. Stephen, Supplementary material for "probabilistic load forecasting for the low voltage network: Forecast fusion and daily peaks", in: Zenodo, 2022, http://dx.doi.org/10.5281/zenodo.7064279.

[19] R.A. Rigby, D.M. Stasinopoulos, Generalized additive models for location, scale and shape, J. R. Stat. Soc. Ser. C. Appl. Stat. 54 (3) (2005) 507–554.

[20] T. Hastie, R. Tibshirani, J.H. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second ed., Springer, 2009, p. 745.

[21] D.M. Stasinopoulos, R.A. Rigby, Generalized additive models for location scale and shape (GAMLSS) in R, J. Stat. Softw. 23 (7) (2007) 1–46, http://dx.doi.org/10.18637/jss.v023.i07.

[22] J. Browell, C. Gilbert, R. Tawn, L. May, Quantile combination for the EEM20 wind power forecasting competition, in: 2020 17th International Conference on the European Energy Market, Vol. 2020-Septe, EEM, IEEE, 2020, pp. 1–6, http://dx.doi.org/10.1109/EEM49802.2020.9221942.

[23] S. Baran, S. Lerch, Combining predictive distributions for the statistical post-processing of ensemble forecasts, Int. J. Forecast. 34 (3) (2018) http://dx.doi.org/10.1016/j.ijforecast.2018.01.005.

[24] K. Kober, G.C. Craig, C. Keil, A. Dörnbrack, Blending a probabilistic nowcasting method with a high-resolution numerical weather prediction ensemble for convective precipitation forecasts, Q. J. R. Meteorol. Soc. 138 (664) (2012) 755–768, http://dx.doi.org/10.1002/qj.939.

[25] C. Capezza, B. Palumbo, Y. Goude, S.N. Wood, M. Fasiolo, Additive stacking for disaggregate electricity demand forecasting, Ann. Appl. Stat. 15 (2) (2021) http://dx.doi.org/10.1214/20-AOAS1417.

[26] T. Gneiting, R. Ranjan, Combining predictive distributions, Electron. J. Stat. 7 (2013) 1747–1782.

[27] R. Ranjan, T. Gneiting, Combining probability forecasts, J. R. Stat. Soc. Ser. B Stat. Methodol. 72 (2010) 71–91.

[28] F. Castanedo, A review of data fusion techniques, Sci. World J. 2013 (2013) http://dx.doi.org/10.1155/2013/704504.

[29] R.J. Abrahart, L. See, Multi-model data fusion for river flow forecasting: An evaluation of six alternative methods based on two contrasting catchments, Hydrol. Earth Syst. Sci. 6 (4) (2002) 655–670, http://dx.doi.org/10.5194/hess-6-655-2002.

[30] H.G. Bergsteinsson, J.K. Møller, P. Nystrup, O.P. Palsson, D. Guericke, H. Madsen, Heat load forecasting using adaptive temporal hierarchies, Appl. Energy 292 (2021) 116872, http://dx.doi.org/10.1016/j.apenergy.2021.116872.

[31] A. Bender, A. Groll, F. Scheipl, A generalized additive model approach to time-to-event analysis, Stat. Model. 18 (3–4) (2018) http://dx.doi.org/10.1177/1471082X17748083.

[32] L. Fahrmeir, Discrete Failure Time Models, Tech. Rep., Collaborative Research Center 386, Discussion Paper 91, 1997, http://dx.doi.org/10.5282/UBM/EPUB.1483.

[33] G. Tutz, H. Binder, Flexible modelling of discrete failure time including time-varying smooth effects, Stat. Med. 23 (15) (2004) http://dx.doi.org/10.1002/sim.1824.

[34] S. Tindemans, G. Strbac, J.R. Schofield, M. Woolf, R. Carmichael, M. Bilton, low carbon London project: Data from the dynamic time-of-use electricity pricing trial, 2013, 2016, http://dx.doi.org/10.5255/UKDA-SN-7857-2.

[35] R. Core Team (2020), R: A language and environment for statistical computing, 2020.

[36] J. Browell, C. Gilbert, ProbCast: Open-source production, evaluation and visualisation of probabilistic forecasts, in: 2020 International Conference on Probabilistic Methods Applied to Power Systems, PMAPS, IEEE, 2020, pp. 1–6, http://dx.doi.org/10.1109/PMAPS47429.2020.9183441.

[37] T. Gneiting, F. Balabdaoui, A.E. Raftery, Probabilistic forecasts, calibration and sharpness, J. R. Stat. Soc. Ser. B Stat. Methodol. 69 (2007) 243–268.

[38] A. Jordan, F. Krüger, S. Lerch, Evaluating probabilistic forecasts with scoring rules, J. Stat. Softw. 90 (12) (2019) 1–37.

[39] D.S. Wilks, Forecast verification, in: Statistical Methods in the Atmospheric Sciences, Elsevier, 2019, pp. 369–483, http://dx.doi.org/10.1016/B978-0-12-815823-4.00009-2.

[40] J.W. Messner, P. Pinson, J. Browell, M.B. Bjerregård, I. Schicker, Evaluation of wind power forecasts: An up-to-date view, Wind Energy (2020).

[41] T.M. Hamill, Reliability diagrams for multicategory probabilistic forecasts, Weather Forecast. 12 (4) (1997) http://dx.doi.org/10.1175/1520-0434(1997)012<0736:RDFMPF>2.0.CO;2.

[42] M. Grabner, Y. Wang, Q. Wen, B. Blažič, V. Štruc, A global modeling approach for load forecasting in distribution networks, 2022, http://dx.doi.org/10.48550/arxiv.2204.00493, arXiv.