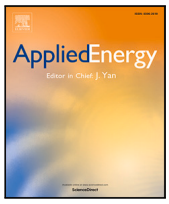
Applied Energy 382 (2025) 125168



Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy





A global probabilistic approach for short-term forecasting of individual households electricity consumption

Lola Botman [a,*], Jesus Lago [b,1], Thijs Becker [c], Koen Vanthournout [c], Bart De Moor [a,2]

[a] KU Leuven, ESAT, Stadius, Belgium
[b] Amazon, Netherlands
[c] Water & Energy Transition (WET) unit, VITO/EnergyVille, Genk, Belgium

ARTICLE INFO

Keywords:
Short-term probabilistic load forecasting
Smart meter
Household consumption

ABSTRACT

Accurate short-term prediction of individual residential load is essential for various applications and low-voltage grid stakeholders. Distribution system operators can utilize these forecasts in grid simulations, operation planning, and to anticipate unusual or varying consumption in certain neighborhoods, which can help to avoid congestions caused by peaks. Furthermore, these predictions can be used to optimize the low-voltage grid for renewable energy sources and potential battery storage. Due to the highly stochastic nature of household consumption, point forecasting is not optimal. We apply state-of-the-art probabilistic methods from other applications, and we propose a novel probabilistic global approach. It is based on a selection of similar normalized household consumption time series, which are then used to compute empirical quantiles. The proposed method has been evaluated on two different open datasets from different countries, and it outperforms state-of-the-art models in both cases. It has also been applied to a private Belgian dataset. In all three cases, the proposed method consistently outperforms the state-of-the-art methods on highly stochastic and changing households, i.e., households experiencing concept drift. This method is scalable, with low computational requirements, requires only seven days of historical data of the target household to make predictions and does not require household-specific or weather information.

1. Introduction

Accurately forecasting the electrical consumption of individual households, also called household load forecasting, is necessary for a wide range of applications. Different applications require different forecasting dimensions, i.e., horizon and spatial aggregation [1]. For example, predicting day-ahead consumption, including potential peaks, allows operators to anticipate and take actions to mitigate congestion.

1.1. Related work

Recently, research on residential consumption forecasting at the household level has gained more attention due to the installation of smart meters. It is however still relatively immature [2] compared to forecasting of loads at higher aggregations levels, such as city or national levels [3,4]. Low voltage and household level forecasting is important for the efficient operation of the energy system, especially with the increasing use of renewable energy sources, electrical vehicles, and heat pumps, which all add complexity and uncertainty to household electrical consumption time series, also called consumption profiles [5]. Short-term household consumption forecasting is also necessary to make the energy network more intelligent [6] and to potentially influence customer behavior towards a reduction of energy usage and congestion avoidance mechanisms. The key difference between household consumption time series and aggregated time series at the substation or city level is the variability of the data. Household consumption time series display higher levels of stochasticity compared to aggregated time series due to the individual inhabitants' behavior.

The majority of research in this field has focused on point forecasting methods, i.e., methods where a single value is predicted per time point in the future, while probabilistic forecasting methods, i.e., methods where a distribution or spread is predicted per time point in the future, are more appropriate and informative due to their ability to provide information about the uncertainty of predictions. The first review paper [5] on probabilistic forecasting included all forecasting dimensions, i.e., from short-term to long-term and also different spatial

* Corresponding author.
E-mail address: lola.botman@esat.kuleuven.be (L. Botman).
[1] Jesus Lago contributed to this work as an outside activity and not as part of his role at Amazon.
[2] Fellow IEEE, IFAC, SIAM.

https://doi.org/10.1016/j.apenergy.2024.125168
Received 16 July 2024; Received in revised form 6 December 2024; Accepted 14 December 2024
Available online 7 January 2025
0306-2619/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

aggregations. However, there is limited research on individual household consumption forecasting, and the authors of [5] only refer to [7]. A recent comprehensive review on low voltage load forecasting [8] also highlights the lack of research at the household level and calls for a shift towards probabilistic methods. More recently, Haben et al. published an entire book on load forecasting at the low voltage level [2]. Household load forecasting is mentioned a few times, mostly to present the multiple challenges due to the dependency on human behavior, unpredictability and peaky nature of the time series. In the 'Advanced and Additional Topics' chapter, a dedicated section further highlights the distinct difficulty and complexity of research in the domain of household load forecasting. There is a strong need for further study of probabilistic forecasting methods, case studies, and evaluation methods at the household level.

Al-Qahtani and Crone [9] provide a comprehensive overview of k-NN applications in electricity demand forecasting, particularly for time series data at the national level. Their work focuses on multivariate k-NN regression, integrating calendar features like holidays and weekends through binary dummy variables. While their study highlights the potential of k-NN in load forecasting, it is a point forecasting approach and focuses on large-scale aggregated data rather than individual households. More recent research has explored hybrid approaches to enhance k-NN's performance in load forecasting. For example, Dong et al. [10] proposed a hybrid model combining k-NN with deep belief networks to improve both the forecasting accuracy and computing efficiency in electrical load interval forecasting. Tajmouati et al. [11] introduced two novel k-NN-based methodologies for time series forecasting, focusing on improving parameter tuning and reducing computational time, further validating the versatility of k-NN in different forecasting contexts.

In the context of residential load forecasting, peak consumption is of significant interest as a high simultaneity of peaks may result in congestion on the low voltage grid. Conventional point forecasting methods and their accompanying loss function (error metric) tend to overlook peaks as they generate substantial errors even when correctly predicted in terms of height and duration, albeit with a small time lag. This phenomenon is referred to as the double peak penalty effect. To address this issue, alternative error metrics can be considered [7]. Another approach is to utilize probabilistic forecasting, which can overcome this limitation; as the predictions now account for the spread of the data, they can more accurately capture variations in the timing of peaks. Using probabilistic methods is the recommended alternative, since they can be assessed objectively using proper scoring functions [2]. Scenario generation is another field of research that allows generating realistic daylong electricity consumption profiles [12].

This paper aims to address certain limitations and challenges in the existing load forecasting literature. One limitation is the lack of comparisons with baselines and/or existing state-of-the-art models [2]. However, conducting such comparisons can be challenging, in cases where the proposed methods are applied to private datasets [13] (or datasets that are hard to access in practice or when the code is not readily accessible [14]). These factors hinder reproducibility and comparisons.

Another limitation is the reliance of most papers on a single dataset, which makes it difficult to draw generalizable conclusions about a model's performance.

Some studies have shown that incorporating additional attributes, such as household-specific information like the number of rooms or heating types, can improve forecasting accuracy [15,16]. However, when there is a need to scale the forecasts to thousands of households in an entire region for example, it is not feasible to obtain detailed information about each individual household.

Weather conditions, primarily temperature, along with humidity and sunshine, are frequently incorporated as exogenous variables in electricity demand forecasting models. Their substantial impact at medium and high voltage aggregation levels is well-documented [17–19]. Moreover, several studies have integrated weather data to predict individual household electricity usage [20]. Two significant observations emerge from the discussions and research on the influence of weather data on model accuracy at the household level. First, most studies focus on electricity demand forecasting using true future weather data, potentially leading to an overestimation of performance enhancements. In practical set-ups, future weather data would need to be forecast, affecting the applicability of these conclusions [19]. Second, even though it has been shown that there is a correlation between household consumption and weather data [21], the most recent literature agrees that there is no causality and weather data as input does not significantly increase forecast performance at the household level [2,19,22]. Another reason to consider excluding weather data is to facilitate the use of models on anonymized data, which may lack specific location information and subsequent relevant weather information. It is also important to note that, when historical consumption data is used to generate forecasts, the impact of weather is often implicitly captured, as past consumption reflects the effects of previous weather conditions. This indirect inclusion may reduce the necessity of explicitly adding weather variables, particularly in short-term forecasts. It is essential to recognize that the impact of weather on household electricity consumption can vary within a dataset, for example, between households with gas heating and those with electric heating [2]. In the future, the influence of weather data on individual household electricity consumption will likely change with the increase in solar panels and heat pumps [2]. This is an interesting area for future research when more data becomes available.

An additional critical challenge in the field of time series forecasting, particularly pertinent to individual household electricity consumption, is the phenomenon of concept drift. Concept drift refers to the change in statistical properties of the variable over time, which can be due to evolving consumer behaviors, technological advancements, or changes in external factors like the covid pandemic. This drift often results in the degradation of forecast accuracy as the models, trained on historical data, fail to adapt to the new data distribution. This is especially problematic in short-term household consumption forecasting, where rapid shifts in user behavior or sudden introduction of new appliances such as electrical vehicles or heat pumps, can significantly alter consumption patterns. The challenge lies in developing forecasting models that are robust to such changes and can adapt over time [23,24].

1.2. Motivation and contributions

To overcome the challenges mentioned in Section 1.1, this paper proposes a new probabilistic approach for short-term forecasting of residential consumption, based on empirical quantiles. Because of the high stochasticity of individual household consumption time series, probabilistic methods are needed to gain insight into the uncertainty of the predictions. The paper contributes to the literature by proposing:

- a novel global probabilistic method that generates accurate short-term forecasts of individual households electrical consumption, consistently outperforming state-of-the-art methods on time series experiencing concept drift;
- a method that generates probabilistic forecasts for households with only seven days of historical data of the target household, leveraging other households' historical data;
- a comprehensive comparison with state-of-the-art probabilistic forecasting methods, not yet applied in short-term household load forecasting, using three different datasets.

The method is evaluated on three distinct datasets. For reproducibility, two of these are publicly available, and the accompanying code is provided [25]. The method presents interesting properties. First, it is global, i.e., it uses one model for all households. This is in contrast with

most methods proposed in the literature which are categorized as local, i.e., one model is trained per household. Grabner et al. [24] also discuss global versus local models in the case of point forecasting. Additionally, the method does not rely on weather or household-specific attributes; it generalizes well, i.e., it outperforms state-of-the-art methods on several datasets; it is scalable, thanks to the low data requirement and low model complexity [26]; and the model is easily understood, unlike black-box models. In practice, to know where a prediction comes from, one can explicitly look at the similar time series, i.e., households that were selected to compute empirical quantiles. This means one can investigate and understand the individual predictions. This paper is an extension of [27].

1.3. Organization of the paper

The paper is organized as follows. Section 2 presents the background knowledge necessary to understand the method, it includes details on empirical quantiles (Section 2.1), the different benchmark models (Section 2.2) and the performance metric (Section 2.3). The different steps of the method are detailed in Section 3, as well as the model parameters. The case study is explained extensively in Section 4, including information on the different datasets (Section 4.2) and the results (Section 4.3). Finally, the outcome and perspectives are discussed in Section 6.

2. Background knowledge

2.1. Empirical quantiles

Quantiles are a fundamental concept in statistics and data analysis. They represent specific points in a distribution that divide the data into equal portions. They provide valuable information on the spread and variability of data. Rob J. Hyndman and Fan [28] examine the calculation of sample quantiles, highlighting variations in existing algorithms, mathematical assumptions and their impact on estimation. Sample quantiles, also called theoretical quantiles are defined based on a theoretical (either given, known or assumed) distribution. Empirical quantiles, on the other hand, are calculated directly from observed data. An empirical cumulative distribution function (CDF) can be constructed by estimating quantiles. The approximation of the CDF is closer to the true CDF, if more observations are available. To estimate a given q-quantile x_q, the dataset is arranged in ascending order, and the q-quantile is the value in the dataset that corresponds to the proportion of observations q that are less than or equal to that value. For example, to compute the median $x_{q=0.5}$, i.e., also called the 0.5-quantile, or the 50th percentile, half of the observations will fall below this value $x_{q=0.5}$ and half will fall above it. Fig. 1, inspired from [2], illustrates an empirical CDF (in dotted blue line) with true CDF (in black) for the standard normal distribution. The empirical CDF is generated using 30 randomly drawn points from the true CDF. The median is also indicated with the light blue line.

Empirical quantiles play a crucial role in understanding the distributional characteristics of a dataset and in estimating percentiles without making assumptions about the underlying distribution. Ruppert and Matteson [29] explore the application of empirical quantiles in financial engineering, specifically in the context of probabilistic forecasting. By using historical data to estimate empirical quantiles, it becomes possible to generate probabilistic forecasts that capture the uncertainty associated with future observations. This approach is particularly useful in financial markets, where volatility and risk management are of great importance. Likewise, a high stochasticity and variability can be observed in individual household data, which makes empirical quantiles relevant. Moreover, there is a parallel objective shared with distribution system operators, namely the efficient management of congestion risk while minimizing costs. Although [30] deals with long-term predictions, it highlights the added value of empirical quantiles in load forecasting. It is worth noting that terminology can vary between different sources and disciplines and that sample quantiles and empirical quantiles are for example used as synonyms in [2,29], but not in [28].

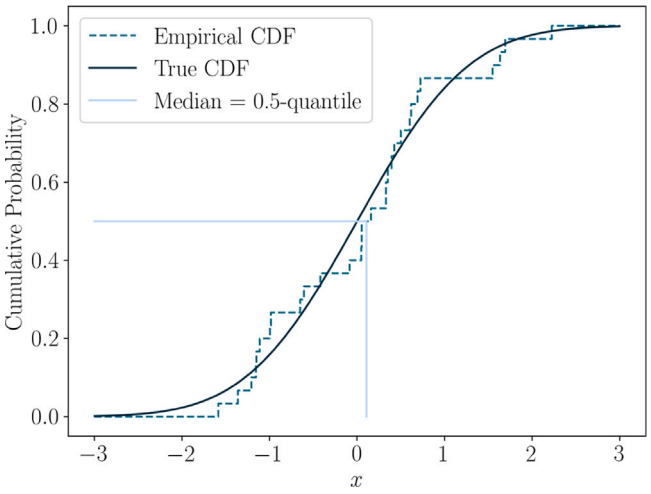


Fig. 1. Empirical CDF (dotted blue) with true CDF (black) for the standard normal distribution. The empirical CDF is generated using 30 randomly drawn points from the true CDF. The median (0.5-quantile) is also illustrated with the light blue line. The figure originally appeared in [2], and has been reproduced in this paper with the addition of the median.

2.2. Benchmark methods

The benchmark models utilized in this study encompass established and state-of-the-art approaches employed in the literature for this specific application or successful in other applications of probabilistic forecasting. The proposed baseline consists of simplistic approaches that rely on naive assumptions and do not require extensive modeling. All described benchmark models rely on same input features, except for the Bernstein Normalizing Flows (Section 2.2.3) that has additional calendar features. Each model generates 100 quantiles at each time step, for a one-day horizon. The train–test split, input and output are consistent across all models and are detailed in the first benchmark model in Section 2.2.1.

2.2.1. Linear quantile regression

A linear quantile regression model predicts quantiles, by optimizing the Pinball loss.[3] There is one model per quantile to be predicted in the future. Quantile regression has the advantage of not making any assumption on the distribution of the dataset. The Python library `scikit-learn` allows for a straightforward implementation. The selected parameters included 100 quantiles, no L1-regularization, i.e., $\alpha = 0$, and the 'highs-ds' solver. The reader is referred to [2,31] for more details on quantile regression and to [32,33] for the application of quantile regression on short-term probabilistic load forecasting. The model input consists of two days of historical consumption data preceding the target day and the corresponding day one week and two weeks ago. These inputs capture recurring patterns in customer behavior. The pre-processing consists only of standard min–max normalization of these consumption values. Furthermore, calendar information is included as input variables, i.e., the hour of the day (hod), the day of the month (dom) and the month of the year (moy), each cyclically encoded, and a binary indicator for weekends versus weekdays (wk). The sine and cosine encoding ensures the algorithm understands the

[3] More details about the Pinball loss are provided in Section 2.3.

cyclical nature of these features, e.g., 23 h00 in the evening is close to 1 h00 in the morning [34,35]. The consumption of a household for day d is a vector noted as $\mathbf{y}_d$. The amount of time steps in one day, T, determines the dimension of $\mathbf{y}_d$. For example, half-hourly time series have 48 time steps in one day, while quarter-hourly time series have 96 time steps in one day. If d is the day to be predicted, the input vector is:

$$\begin{bmatrix}\mathbf{y}_{d-14}, \mathbf{y}_{d-7}, \mathbf{y}_{d-2}, \mathbf{y}_{d-1}, \\ \sin(\text{hod}), \cos(\text{hod}), \\ \sin(\text{dom}), \cos(\text{dom}), \\ \sin(\text{doy}), \cos(\text{doy}), \text{wk}\end{bmatrix}_{4T+7}. \tag{1}$$

The model produces 100 quantiles for each time step in the one-day-ahead forecast horizon, resulting in an output of size $100{\times}T$. The data is split sequentially in time, with the first 60% allocated for training, the next 20% for validation, and the final 20% for testing. It is important to note that depending on the dataset length, the test set, i.e. the last 20% may or may not include months present in the training set. For instance, if the dataset spans only one year, the training and validation sets will cover data from January to mid-October, while the test set will cover the remaining part of October, as well as November and December. However, if the dataset extends beyond one year, the test set will include months that are already present in the training set.

2.2.2. Quantile Regression Averaging (QRA)

Quantile Regression Averaging is relatively new to the field of probabilistic load forecasting. Nevertheless, QRA is recognized as state-of-the-art probabilistic machine learning in electricity price forecasting applications [36]. Liu et al. [37], applied it on aggregated load data. To the best of the authors' knowledge, it is applied to individual household load data for the first time in this paper. QRA starts by generating multiple point forecasts from different models, called "sister forecasts". In this work, the sister forecasts are built using four well-known point forecasting methods: a feed-forward neural network [27], a lasso regression [38,39], an XGBoost model [40,41], and a naive persistence model, i.e., the predicted consumption for a given day is the consumption observed on the previous day. The point forecasting models produce a single value for each time step in the one-day ahead horizon. QRA constructs quantiles by combining the ensemble of sister forecasts. This method employs quantile regression to generate 100 quantiles per time step, for a one-day horizon. The four point forecasts serve as inputs for the quantile regression process. This averaging step aggregates the individual predictions, resulting in an improved overall probabilistic forecast that captures a wider range of potential outcomes than a single point forecast. QRA has the advantage of leveraging the existing literature on point forecasting and does not rely on high-quality expert forecasts [37]. The same inputs, outputs, and data split as described in Section 2.2.1 are considered. Regarding hyperparameters, QRA uses the `highs − ds` solver. In the FFNN point forecast models, the learning rate is set to 0.001, batch size to 64, and the models are trained for up to 1000 epochs, with early stopping used for regularization. In the LASSO point forecast model, the regularization parameter alpha is set to 0.001. In the XGBoost point forecast model, the number of trees is set to 1000, and the squared error is used as the objective function. The remaining parameters are set to default.

2.2.3. Bernstein Normalizing Flows (BNF)

This model uses Bernstein normalizing flows to transform a simple distribution into a more complex distribution. A neural network is trained to predict the parameters of the flows. These parameters yield the conditional probability distributions for all the time steps, one day ahead. Seven days of data prior to the predicted day are used as input and there are five additional calendar inputs: the sine and cosine encoding of the day of the year, the sine and cosine encoding of the day of the week, and a binary indicator for the holidays. This approach differs from the other benchmark models, which use 14 days, as the model is implemented according to the description in the original paper. This may present a potential disadvantage. The authors claim that their method outperforms a simple Gaussian Model, a Gaussian Mixture Model and a Quantile Regression. The reader is referred to the full papers and implementation for details [42–44]. As this approach outputs distribution functions per time step, we computed quantiles from these distributions and then applied our implementation of the discretized Continuous Ranked Probability Score (CRPS) on the quantiles (detailed in Section 2.3). Note that the BNF method results presented in this paper are slightly different from the results in [42,44] as we have re-run them locally, with our (not normalized) CRPS implementation, to allow for exact comparison with the other methods.

2.2.4. Feedforward Neural Network (FFNN) & empirical quantiles

This method consists of three main modules applied to each household consumption time series individually: the preprocessing, the point forecasting and the calculation of empirical quantiles. Each time series is preprocessed individually; a feedforward neural network point forecasting model is trained on the training set; this model is then applied to the validation set. The point forecast errors on the validation set are computed and grouped per time step, i.e., the first group consists of all prediction errors made at the first time step (00 h00), the second group consists of all errors made at the second time step (00 h30), and so on. There is one set of errors for each time step of the day. Empirical quantiles are then computed based on these groups and finally, the point forecasting model is applied on the test set and the empirical quantiles are added to the point forecast to generate probabilistic forecasts. This approach is detailed in a previous paper written by the authors, and the code is also made available [25,27]. The same inputs, outputs, and data split as described in Section 2.2.1 are considered. The Adam optimizer with default parameters from `Keras` is used for the training, i.e., learning rate = 0.001, beta 1 = 0.9, beta 2 = 0.999, epsilon = $1e{-}07$. Early stopping is used during the training, to avoid overfitting.

2.2.5. Baselines

Two different baseline methods are proposed in this work. The first one is the same as in [27]. One can assume that, in the absence of mathematical formulations or complex modeling, the most reasonable hypothesis is that the current day's consumption will resemble that of the previous day, commonly known as a persistence model in point forecasting. In line with this assumption, we propose a probabilistic baseline that shares the same grounding. In Section 2.2.4, we introduced the FFNN + Empirical Quantiles model. Building upon this model, the baseline utilizes the persistence model for point forecasting instead of the FFNN, and subsequently applies the empirical quantiles' step. The second baseline, named historical sampling, relies on the same assumption, but instead of going through a point forecasting step, the empirical quantiles are computed directly on the historical consumption, e.g., the predicted quantiles of the consumption of a Monday at 15:00 are computed using the consumption of all previous Mondays at 15.00.

2.3. Performance metrics

In the context of point forecasting, where a prediction of a specific value at a given time step is made, the evaluation of the prediction's accuracy is commonly quantified using error measures such as the Mean Absolute Error (MAE) and the Mean Squared Error (MSE), where the absolute (or squared) error is computed at each time step and then averaged:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| = \frac{1}{n}\sum_{i=1}^{n}\left|e_i\right|, \tag{2}$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2, \tag{3}$$

where y_i are the true consumption values, $\hat{y}_i$ are the predicted consumption values and n is the number of samples. In the case of probabilistic forecasting, where a distribution of potential future outcomes is predicted directly, the evaluation of the prediction's accuracy could be quantified by taking the mean or the median of the predicted distribution, and computing the Mean Absolute Error (MAE) and Mean Squared Error (MSE) with respect to the true values [42,45]. However, these conventional metrics do not fully capture the probabilistic information included in the prediction. Therefore, it is necessary to use a metric that is better suited for probabilistic forecasts, such as the Continuous Ranked Probability Score (CRPS), which takes into account both the calibration (accuracy) and sharpness (spread) of the predicted distribution [46–48]. The CRPS is calculated individually for each predicted time step and then averaged over the entire horizon. The metric for one measurement can be computed as

$$\text{CRPS}\left(\hat{F}, y\right) = \int_{-\infty}^{\infty}\left(\hat{F}(x) - \mathbb{1}\{y \le x\}\right)^2 \mathrm{d}x \tag{4}$$

where $\hat{F}$ is the predicted cumulative distribution function, y is the observed consumption value, and $\mathbb{1}$ is the indicator (also called Heaviside) function, i.e. takes value 1 when $y \le x$ and 0 otherwise (in blue line on Fig. 2). The CRPS is a commonly used metric for probabilistic forecasting and can be intuitively interpreted as a sum of areas[4] between the predicted CDF and the actual value. It should be noted that the CRPS entails a more intricate calculation than solely computing the area between the predicted CDF and the observed value, as illustrated in details in Fig. 2 [50]. However, the CRPS evaluates the continuous CDF, while our proposed method generates quantiles. It is thus important to consider another version of the CRPS [48], which uses the quantiles, in order to evaluate the performance of the proposed method:

$$\text{CRPS}\left(\hat{F}, y\right) = \sum_{q\in[0,1]}\left\{\text{Pinball}\left(\hat{Q}(q), y, q\right)\right\}, \tag{5}$$

where

$$\text{Pinball}\left(\hat{Q}(q), y, q\right) = \begin{cases}(1-q)\left(\hat{Q}(q) - y\right), & \text{for } y < \hat{Q}(q),\\ q\left(y - \hat{Q}(q)\right), & \text{for } y \ge \hat{Q}(q),\end{cases} \tag{6}$$

and $\hat{Q}(q)$ is estimated the q-quantile of the consumption and y is the actually observed consumption. Our models generate 100 quantiles and the sum is thus computed for $q = [0, 0.01, 0.02, \ldots, 0.99, 1]$.

The Pinball loss is also called the quantile loss, and measures the accuracy of one quantile. It is an asymmetric loss function that takes the difference between the observed value and the predicted quantile and assigns a different weight based on the sign of this difference. Indeed, the q-quantile $\hat{Q}(q)$ is accurately predicted when a proportion q of the observation are below $\hat{Q}(q)$ [2]. The Pinball loss score, also called the quantile score, is obtained by averaging over multiple quantiles, and multiple time steps. A lower pinball score indicates a better forecast. The discretized CRPS can also be interpreted visually, as illustrated in Fig. 2(b), it is a sum of overlapping rectangular area, shaded in green. Furthermore, the CRPS is a generalization of the MAE, to which it reduces if $\hat{F}$ is a point forecast [51].

Note that we compute error metrics on the unnormalized values. Unnormalized error metrics offer valuable insights into the absolute performance of forecasting models, providing distribution system operators with the true error values relevant to their applications. Additionally, absolute metrics allow us to identify challenging time series.

[4] Note that several references illustrate the CRPS as the exact area between the indicator function and the predicted CDF. While this intuition behind the CRPS is correct, the representation is not entirely accurate [2,49].

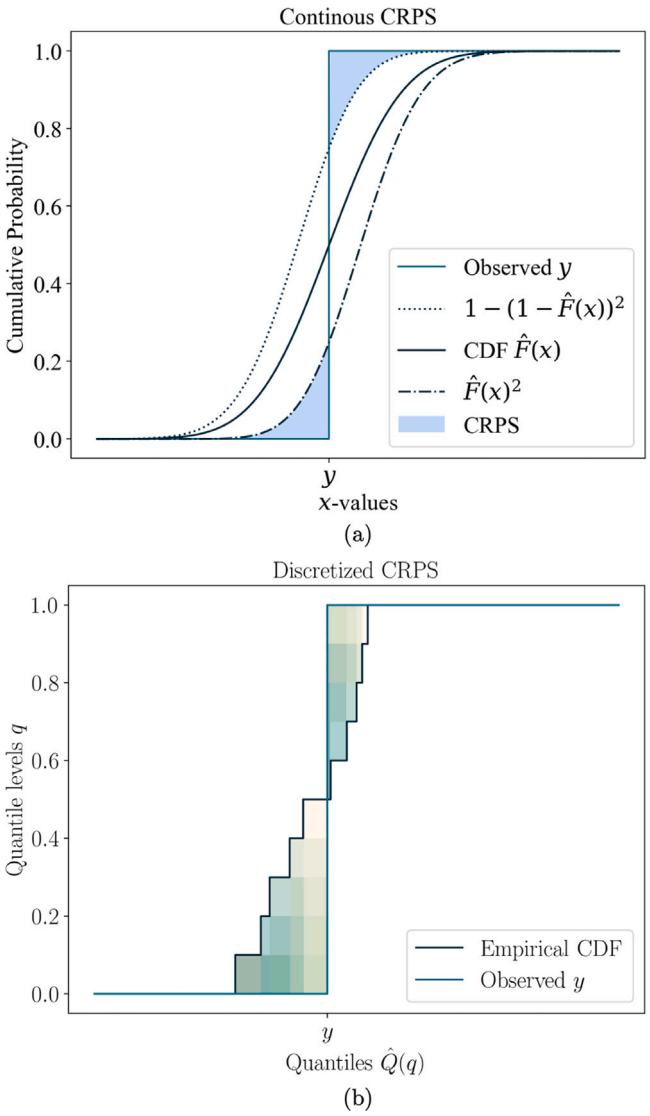


Fig. 2. (a) Visual representation of the continuous CRPS computation between predicted cumulative distribution (black) and observed distribution (blue) [50]. The CRPS is the light blue colored area. (b) Visual representation of the discretized CRPS computation between empirical cumulative distribution (black) and observed distribution (blue). The CRPS is calculated as the sum of all the overlapping rectangular areas shaded in green. It is important to note that for the sake of clarity in this illustration, only ten quantiles are shown, while the proposed method computes 100 quantiles.

3. Methodology: Global approach

The method generates day-ahead probabilistic predictions for individual household consumption time series, where 100 quantiles are provided for each time step. The data cleaning is presented in Section 3.1. The approach consists of three sequential steps applied on each time series: normalization, selection of similar households and prediction. The steps are detailed in Sections 3.2–3.4. The method can easily be adapted to a specific dataset thanks to the modular approach. The same data split as described in Section 2.2.1 is considered, where the first sequential 80% of the dataset is allocated for training, and the remaining 20% is reserved for testing.

3.1. Data cleaning

The data cleaning and pre-processing is highly dependent on the dataset used. In each case it is kept to the bare minimum to allow for

reproducibility on other datasets. Only residential customers are used, as it is of specific interest to this research. Isolated missing values, i.e., one missing time step between two non-missing time steps, are linearly interpolated and incomplete profiles, i.e., time series where several weeks or months are missing, are discarded. The amount of time steps in one day depends on the resolution of the dataset, e.g., half-hourly time series have 48 time steps in one day, while quarter-hourly time series have 96 time steps in one day. The number of time steps in one day is denoted by T, the daily consumption of a household j for day d is noted as:

$$\mathbf{y}_d^j = \left[y_{d_1}^j, y_{d_2}^j, \dots, y_{d_T}^j\right] \in \mathbb{R}^T, \tag{7}$$

where $y_{d_t}^j$ is the quarter-hourly or half-hourly electricity consumption of household j, at time step t, during day d.

3.2. Normalization

The goal of the normalization step is to enable the comparison of different time series, i.e. household profiles, with each other, to find similar patterns and use these later on to make predictions. Different types of normalization are used, depending on the properties of the dataset. In this study, the normalization process involves scaling the data between 0 and 1 in the case of positive values only, or between −1 and 1 for datasets with negative values. Negative values correspond to an excess electricity production of solar panels injected on the grid by households. The scaling is based on the maximum absolute value of the training set,[5] ensuring that zero values remain unchanged throughout the normalization procedure, and that the test set is unused.[6]

- Profile normalization[7]: the time series is normalized between 0 and 1, or between −1 and 1 for datasets with negative values, according to the maximum absolute value of the training set. The profile normalization allows to compare household profiles with similar patterns, but different absolute values. For a quarter- or half-hourly value y_t^j, of household j, at time step t, the "profile normalized" value, indicated by the subscript p, is

$$r_{p,t}^j = \frac{y_t^j}{\left|y_p^j\right|_{\max}}. \tag{8}$$

  where $\left|y_p^j\right|_{\max}$ is the maximum absolute consumption in the training set of household j.
- Week normalization: Each set of seven days is normalized between 0 and 1, or between −1 and 1 for datasets with negative values. This allows for comparison between household profiles and within a household profile. We assume that weekly patterns might be present through the year, which might go unnoticed with the profile normalization. For a quarter- or half-hourly value y_t^j of a given week w of household j, at time step t, the "week normalized" value, indicated by the subscript w, is

$$r_{w,t}^j = \frac{y_t^j}{\left|y_w^j\right|_{\max}}. \tag{9}$$

  where $\left|y_w^j\right|_{\max}$ is the maximum absolute consumption in the week w of household j.
- Day normalization: Each day is normalized between 0 and 1, or between −1 and 1 for datasets with negative values. As in the week normalization, the idea is to look for similar days between household profiles and within a profile with a finer granularity than the week normalization. For a quarter- or half-hourly value y_t^j, of a given day d, at time step t, of household j, the "day normalized" value, indicated by the subscript b, is

$$r_{b,t}^j = \frac{y_t^j}{\left|y_b^j\right|_{\max}}. \tag{10}$$

  where $\left|y_b^j\right|_{\max}$ is the maximum absolute consumption of in the day b of household j.

A normalized full day d, irrespective of the normalization used, is denoted by $\mathbf{r}_d^j$, in coherence with Eq. (7):

$$\mathbf{r}_d^j = \left[r_{d_1}^j, r_{d_2}^j, \dots, r_{d_T}^j\right] \in \mathbb{R}^T. \tag{11}$$

If relevant, the type of normalization is specified by the subscript: $\mathbf{r}_{p,d}^j$ for a "profile normalized" day, $\mathbf{r}_{w,d}^j$ for a "week normalized" day, and $\mathbf{r}_{b,d}^j$ for a "day normalized" day. In all cases, the previous maximum (from the training set, the previous week, or the previous day) is used to unnormalize. An alternative approach would be to train a model to forecast the next day's maximum for unnormalization, which is planned for future work.

3.3. Selection of similar households

This section provides a detailed explanation of the procedure to select similar households, regardless of the normalization used. The general idea is to examine the consumption patterns of a given household during the seven days prior to the day to be predicted and to compare these seven days with all seven-day periods of all households in the historical dataset. Utilizing the k-nearest neighbors algorithm, a number of households that exhibit the closest resemblance (along with their corresponding seven-day periods) are selected. Our method then uses the consumption data of these selected similar households, as detailed in Section 3.4 to make predictions, and thus only requires minimal historical data of the household to be predicted, i.e., the seven days prior to the target day, leveraging other households' historical data.

First, the set $\mathcal{S}_d^j$ is formed, consisting of the seven normalized consumption days of household j, prior to the target day d, where normalized consumption values are denoted as $\mathbf{r}_d^j$:

$$\mathcal{S}_d^j = \left[\mathbf{r}_{d-1}^j, \mathbf{r}_{d-2}^j, \dots, \mathbf{r}_{d-7}^j\right]^\intercal \in \mathbb{R}^{7\times T}. \tag{12}$$

Then, a collection $\mathcal{P}$ is built, using a rolling window of seven days, with a one-day increment (and thus a six-day overlap), over each household profile up to and excluding day d. The collection $\mathcal{P}$ is noted as:

$$\mathcal{P} = \{\mathcal{S}_8^1, \dots, \mathcal{S}_d^1, \dots, \mathcal{S}_8^j, \dots, \mathcal{S}_d^j, \dots, \mathcal{S}_8^n, \dots, \mathcal{S}_d^n\}. \tag{13}$$

The set of seven days prior to the day to be predicted $\mathcal{S}_d^j$ is then compared to the collection $\mathcal{P}$, and the k most similar sets are selected using the k-nearest neighbors algorithm [52], employing the Euclidean distance metric. Note that all of the historical consumption of

[5] It should be noted that the normalized test data might be outside of the [−1,1] interval if there is a value in the test set larger than the maximum value of the training set.

[6] We also implemented and applied a robust scaler, based on the median and the interquartile range [2], which recommended to avoid the effect of outliers. However, this did not yield better results and was not included in the paper.

[7] We also implemented profile normalization, after removing the unusual high peaks: first, the maximum m% of the training set absolute values are replaced by the maximum of the remaining $100 - m$% of the training set and then the "no peak" profile is normalized between 0 and 1, or between −1 and 1 for datasets with negative values. The idea to remove the very high peak values is to allow for comparison, without taking into account households' potentially unusual and unrepresentative behavior. However, we did not notice any improvement and thus did not include it in the results.

household j is included in $\mathcal{P}$ and used to make predictions. One could also choose to exclude household j from the collection $\mathcal{P}$, assuming that only seven days of historical consumption of the household to be predicted are available. The value of the parameter k, and the normalization type are varied and the best performing combination is selected, as detailed in Section 4.4.

3.4. Prediction

One hundred empirical quantiles [28] are computed per time step to be predicted, based on the last day of the selected k-nearest neighbors. The method is based on successful examples of use in weather and price forecasting techniques [48,53]. For each predicted day d, of household j, the output of the model is a matrix of T vectors, i.e., one for each of the T time steps:

$$\hat{Q}_d^j = \left[\hat{\boldsymbol{Q}}_{\boldsymbol{d_1}}^{\boldsymbol{j}}, \hat{\boldsymbol{Q}}_{\boldsymbol{d_2}}^{\boldsymbol{j}}, \ldots, \hat{\boldsymbol{Q}}_{\boldsymbol{d_T}}^{\boldsymbol{j}}\right] \in \mathbb{R}^{100\times T}, \tag{14}$$

and each vector $\hat{\boldsymbol{Q}}_{\boldsymbol{d_t}}^{\boldsymbol{j}}$ contains the 100 predicted quantiles of household j, on day d, at time step t:

$$\hat{\boldsymbol{Q}}_{\boldsymbol{d_t}}^{\boldsymbol{j}} = \left[\hat{Q}_{d_t}^j(0.01), \hat{Q}_{d_t}^j(0.02), \ldots, \hat{Q}_{d_t}^j(1)\right]^{\mathsf{T}} \in \mathbb{R}^{100}. \tag{15}$$

Details on the computation of empirical quantiles are provided in Section 2.1. In practice, the `nanquantile()` function from the `numpy` package is used. The implementation is available in the open code repository [25].

4. Case study

4.1. Problem statement

The proposed method produces household electricity consumption forecasts, in a probabilistic manner, i.e., the forecasts consist of 100 quantiles per time step for a daily horizon. Only seven days of historical consumption of the target household is necessary and no other exogenous data is required such as household's attribute and weather information. The other household profiles in the dataset are leveraged to make predictions.

4.2. Datasets

4.2.1. Irish open dataset

The first dataset used in this study is sourced from the Commission for Energy Regulation (CER) of Ireland and is publicly available on the Irish Social Science Data Archive (ISSDA) website [54]. The data was collected as part of the Smart Metering Project - Electricity Customer Behaviour Trial, and comprises half-hourly ($T = 48$) electrical consumption measurements (in kWh) of over 5000 Irish homes and businesses, spanning from July 2009 to December 2010. To ensure consistency in the benchmarking process, the preprocessing pipeline outlined in a previously published study [42,43] is employed. As described in the cited study, non-residential customers are removed, and only complete records are retained, as the focus of the model is on the stochastic behavior of residential customers. A subset of the remaining time series is then selected randomly, specifically 10% of the dataset, equivalent to 363 customers.

4.2.2. British open dataset

The second dataset is made available by the UK Power Networks [55]. During a four-year innovation project, called Low Carbon London, they recorded the half-hourly ($T = 48$) electricity consumption of 5567 London households between November 2011 and February 2014. A representative sample of the Greater London population was recruited as participants. The data set includes two groups of customers: a sub-group of approximately 1100 individuals who were exposed to Dynamic Time of Use (dToU) energy prices during the 2013 calendar year, and a larger group of approximately 4500 individuals whose energy consumption was not subject to the dToU tariff. The dToU prices were communicated to the sub-group on a day-ahead basis through the Smart Meter In-Home Display or via text message. The remaining customers were on a flat rate tariff. We selected a random sample of 363 customers out of the flat rate tariff customers, in order to have the same amount of time series as in the previous dataset. Very simple preprocessing is applied: removing duplicate data and linear interpolation of missing values (only 0.14% of the data is missing).

4.2.3. Private belgian dataset

This private dataset is provided by Fluvius, the Flemish distribution system operator. They collected the electricity consumption of more than 3000 households in Flanders, for a year, between 2014 and 2016. For this application, 363 one-year time series are randomly selected, as in the two previous datasets. Each time series consists of the quarter-hourly ($T = 96$) electrical consumption (in kWh). It should be noted that before being usable, this dataset required the most cleaning, since it is used for the first time. The reader is referred to [56] for more details on data cleaning and data quality of a sample of 100 household profiles, available publicly [57] from the same distribution system operator.

4.3. Results & discussion

The performances of the proposed approach and benchmarks on the CER, LCL, and Fluvius datasets are summarized in Table 1a, b, and c, respectively. Additionally, a visual representation of the tables can be found in Fig. 3. It is important to note that the magnitude of the CRPS values is influenced by the scale of the respective dataset and that a smaller value of CRPS indicates a better performance of the algorithm. The percentage improvement compared to the baseline is also given. A detailed analysis of the statistical significance of the performance differences between the global approach and alternative methods, including an assessment of the Diebold–Mariano test results, is provided in Appendix. For the CER dataset (Table 1a), our proposed global approach achieved a CRPS of 0.231002, representing a percentage improvement of 31.82% compared to the baseline. This outperformed all other methods such as historical sampling and quantile regression, which showed slightly lower performances. Notably, the proposed approach demonstrated substantial superiority over Bernstein Normalizing Flows and the Feedforward Neural Network + Empirical Quantiles, with percentage improvements of 27.55% and 25.89% respectively.

In the case of the LCL dataset, Table 1b, our proposed global approach achieved a CRPS of 0.091197, resulting in a percentage improvement of 27.23% compared to the baseline. The method outperformed all other approaches such as quantile regression averaging, advanced quantile regression averaging, and historical sampling, which achieved CRPS values of 0.096869, 0.096869, and 0.098104, respectively. Additionally, the proposed approach demonstrated better performance compared to quantile regression and Feedforward Neural Network + Empirical Quantiles, with percentage improvements of 21.34% and 20.25%, respectively.

Lastly, on the Fluvius dataset, Table 1c, the proposed global approach achieved a CRPS of 0.030159, demonstrating a significant percentage improvement of 29.86% compared to the naive baseline. This method also outperformed the Feedforward Neural Network + Empirical Quantiles. Other approaches, including advanced quantile regression averaging, quantile regression averaging, and historical sampling showed superior performance.

The primary distinction between the Fluvius dataset and the CER and LCL datasets lies in their time resolution. To explore the impact of this difference, the Fluvius dataset was resampled at a half-hourly resolution, aligning with the resolution of the other datasets. The results

are presented in Table 1d. Contrary to our initial hypothesis that the increased frequency and extended prediction horizon (in number of time steps) might affect the outcomes, the experiment revealed that the performance ranking of the methods remains consistent across different resolutions. In addition to time resolution, other distinguishing factors between the Fluvius dataset and the others were identified. The Fluvius dataset exhibits higher skew and kurtosis, as detailed in Table 2, indicating substantial deviation from a normal distribution in both the original quarter-hourly and aggregated half-hourly data. Note that the method proposed does not assume normality. Unique to the Fluvius dataset is the presence of negative values, reflecting surplus solar energy injection into the grid. Furthermore, the Fluvius dataset comprises only one year of data, compared to the more extensive data span in the CER and LCL datasets. Additionally, the data quality of the Fluvius dataset is acknowledged to be lower than that of the CER and LCL datasets. These factors – the higher skew and kurtosis, the presence of negative values, the shorter data span, and the lower data quality – might contribute to the observed performance disparities between the Fluvius dataset and the CER and LCL datasets. Recognizing these challenges is crucial for guiding future research and method development in short-term forecasting.

Fig. 4 displays three scatter plots illustrating the individual Continuous Ranked Probability Score (CRPS) values per household, comparing the global approach with the second (or first) best approach for each dataset. In the first sub-figure, the proposed global approach is compared with the historical sampling method using the CER dataset. The second sub-figure presents the comparison between the global approach and the classic quantile regression averaging technique applied to the LCL dataset. In the third sub-figure, the global approach is compared against the advanced quantile regression approach using the Fluvius dataset. The diagonal line within each scatter plot represents the CRPS values at which both methods yield equal performance. As mentioned in Section 2.3, the lower the CRPS, the better the prediction. The data points in the top triangle of the plot, in a dark color, represent the households where the global approach outperforms the alternative method. The data points in the bottom triangle, represented in a light color, indicate the households where the alternative method surpasses the global approach. Notably, across all three datasets, the majority of values closely align with the diagonal line, with no outliers observed in the light-colored data points. However, the dark-colored data points exhibits significant outliers. Consequently, we conclude that the two methods either perform equally well or exhibit comparable performance. However, in cases where there is a substantial difference in performance, the proposed global approach consistently outperforms the alternative method. The global approach manages to avoid outliers with very faulty forecast, and hence is more robust.

The analysis and visualization of specific households in the scatter plots reveals valuable insights regarding the performance comparison between the global approach and the alternative method. Fig. 5 provides a selected representation of households to support the argument. The observations can be extended to encompass a larger number of household profiles, which cannot be included due to space constraints. Differences in the CRPS were computed on a per-household basis, indicating whether the global approach outperformed (positive difference) or the alternative method outperformed (negative difference). Several observations were made based on these differences. Firstly, across all datasets (CER, LCL, and Fluvius), households with the largest positive differences were characterized by significant changes in consumption between the training and test sets, i.e., household profiles experiencing concept drift. A similar conclusion is made in [24], about a global point forecasting method, compared to local point forecasting benchmarks. Secondly, it was noted that the average negative difference was smaller than the average positive difference, indicating that when the alternative method outperformed the global approach, the difference was not as pronounced, whereas when the global approach outperformed the alternative method, the difference was more substantial. Lastly, there

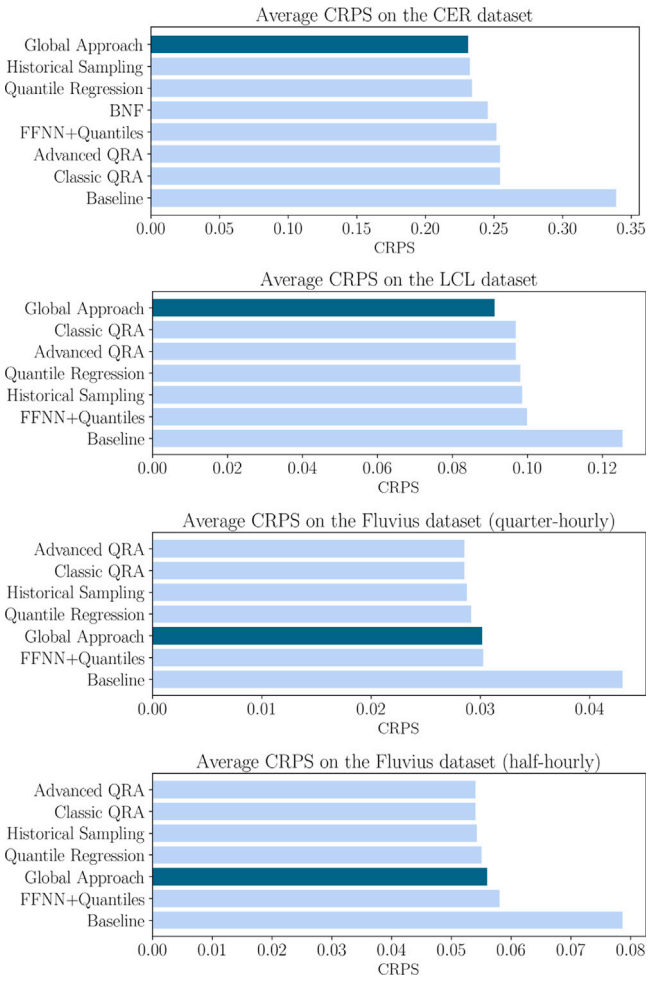


Fig. 3. A visual comparison of the benchmark methods. These are the same results as in Table 1a, b and c, i.e., a naive baseline, historical sampling, a feed-forward neural network enhanced with empirical quantiles [27], quantile regression averaging, quantile regression and Bernstein polynomial flows [44] on the Commission for Energy Regulation (CER) of Ireland open dataset, the London Low Carbon (LCL) open dataset and the Belgian Fluvius private dataset, quarter-hourly and half-hourly. CRPS stands for Continuous Ranked Probability Score, smaller values indicate a higher accuracy of the model. The proposed global approach is highlighted in a darker color.

was no discernible pattern observed among households with the largest negative differences.

The interpretability of the proposed method is enhanced by its reliance on transparent, data-driven processes. Unlike black-box models, where predictions can be challenging to interpret, our approach enables a direct understanding of how individual forecasts are generated. Specifically, the selection of similar historical time series allows users to trace each prediction back to the contributing data from other households. This traceability provides a clear link between the input data and the output forecast, making it possible to understand the patterns that led to the predicted values.

4.4. Hyperparameter sensitivity study of the proposed method

Table 3 presents a comparison of the proposed global approach's accuracy on Commission for Energy Regulation (CER) of Ireland dataset, London Low Carbon (LLC) dataset, and Fluvius dataset. The evaluation focuses on varying the two method's parameters: the normalization type and the number of nearest neighbors. Note that the normalization type and parameter k method were not optimized in a formal tuning process. Various normalization type and values of k were manually

Table 1
A comparison of the benchmark methods, i.e., a naive baseline, historical sampling, a feedforward neural network enhanced with empirical quantiles [27], quantile regression averaging, quantile regression and Bernstein polynomial flows [44] on (a) the Commission for Energy Regulation (CER) of Ireland open dataset, (b) London Low Carbon (LCL) open dataset and the Belgian Fluvius private dataset with (c) quarter-hourly resolution (d) half-hourly resolution. The method proposed in this paper is highlighted, respectively with optimal hyper parameters k and normalization type. CRPS stands for Continuous Ranked Probability Score, smaller values indicate a higher accuracy of the model.

(a) Results on the Commission for Energy Regulation (CER) of Ireland open dataset.

Method	CRPS (CER)	Percentage improvement
Our proposed global approach ($k = 100$, p-norm)	0.231002	31.82%
Historical sampling	0.232272	31.44%
Quantile regression	0.233792	30.99%
Bernstein normalizing flows	0.245456	27.55%
FFNN + Empirical quantiles	0.251072	25.89%
Advanced QRA	0.254295	24.95%
QRA	0.254306	24.94%
Baseline	0.338826	

(b) Results on the Low Carbon London (LCL) open dataset.

Method	CRPS (LCL)	Percentage improvement
Our proposed global approach ($k = 100$, w-norm)	0.091197	27.23%
QRA	0.096869	22.70%
Advanced QRA	0.096869	22.70%
Historical sampling	0.098104	21.72%
Quantile regression	0.098567	21.34%
FFNN + Empirical quantiles	0.099935	20.25%
Baseline	0.125323	

(c) Results on the Belgian Fluvius private dataset with quarter-hourly resolution.

Method	CRPS (Fluvius)	Percentage improvement
Advanced QRA	0.028564	33.57%
QRA	0.028564	33.57%
Historical sampling	0.028786	33.06%
Quantile regression	0.029166	32.17%
Our proposed global approach ($k = 200$, w-norm)	0.030159	29.86%
FFNN + Empirical quantiles	0.030266	29.61%
Naive baseline	0.043003	

(d) Results on the Belgian Fluvius private dataset with half-hourly resolution.

Method	CRPS (Fluvius)	Percentage improvement
Advanced QRA	0.054011	31.29%
QRA	0.054015	31.28%
Historical sampling	0.054264	30.96%
Quantile regression	0.055038	29.98%
Our proposed global approach	0.056001	28.75%
FFNN + Empirical quantiles	0.058060	26.13%
Naive baseline	0.078603	

Table 2
This statistical comparison highlights distinct differences in the Fluvius dataset compared to the other two datasets, observable at both resolutions. Notably, the Fluvius dataset uniquely features negative values, attributed to surplus electricity from solar panels injected into the network. Furthermore, this dataset exhibits substantial skewness and kurtosis. Finally, it is the dataset with the least number of days available. Note that the method proposed does not assume normality.

	Fluvius dataset (quarter-hourly)	Fluvius dataset (half-hourly)	LCL dataset	CER dataset
Number of time series	363	363	363	363
Number of days	365	365	430	535
Resolution	15 min	30 min	30 min	30 min
Missing data (%)	0.121	'0'	0.139	0
Mean	0.059	0.118	0.232	0.515
Std	0.102	0.192	0.334	0.717
Minimum	−0.75	−1.29	0	0
Maximum	3.01	5.95	8.87	19.17
Skew	6.15	6.24	4.59	3.79
Kurtosis	64.41	70.51	34.32	25.33

explored using a grid search approach. Further research could include a formal optimization of the two method parameters. The results reveal that the optimal parameter settings for achieving the best performance differ across datasets. This suggests that the inherent dynamics of each dataset influence the effectiveness of the approach. Furthermore, an analysis of the maximum improvement observed between the "worst" and "best" parameter combinations indicates an improvement of up to 8%. This highlights the non-negligible impact of selecting appropriate parameters on the model's performance.

5. Limitations and areas for further research

While the proposed global model is designed for scalability and reduced computational requirements, an advantage highlighted and proven in this paper, it can present drawbacks in certain cases. First, in scenarios where less household time series are available or where household consumption patterns are highly unique, the global approach may underperform compared to the local models, which train a separate model per household. Analyzing the impact of the number of available time series data on the performance of the model is an interesting research avenue.

A related limitation involves the minimum historical data requirement of seven days. The model is designed to work with only seven days of consumption data from the target household by leveraging historical data from other households to make predictions. This approach is particularly useful when only limited data is available for a new household. However, while the model can function with just seven days of data from the target household, we have not yet specifically tested this scenario. Using only a minimal amount of data may impact forecast quality. We anticipate that model performance would improve as more historical data becomes available. Future work could examine

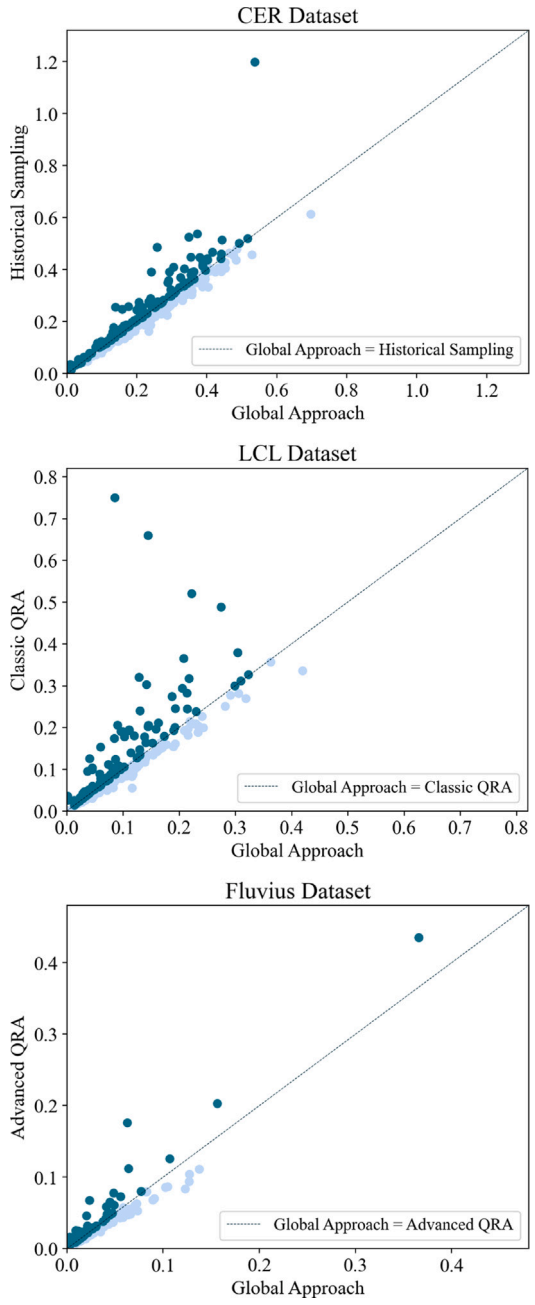


Fig. 4. Each scatter plots compare the global approach with the second (or first) best approach for each dataset. Each dot represents one household, the x-value is the individual household's CRPS when the predictions for that household are made using the global approach, the y-value is the individual household's CRPS when the predictions for that household are made using the alternative second (or first) best approach. The first sub-figure compares the proposed global approach with historical sampling on the CER dataset. The second sub-figure shows the comparison between the global approach and classic quantile regression averaging on the LCL dataset. The third sub-figure compares the global approach with advanced quantile regression on the Fluvius dataset. The diagonal line in each scatter plot represents equal performance for both methods. The dark blue colored data points in the top triangle indicate households where the global approach outperforms the alternative method, while light blue colored data points in the bottom triangle represent households where the alternative method performs better. Notably, most values align closely with the diagonal line across all datasets, with no outliers in the light-colored data points. However, significant outliers exist in the dark-colored data points. These findings suggest that the two methods generally perform equally well or comparably, but the proposed global approach consistently outperforms the alternative method in cases of outlier households.

Table 3
A comparison of the proposed global approach accuracy on the three datasets when varying the normalization type and the number of nearest neighbors k. CRPS stands for Continuous Ranked Probability Score, smaller values indicate a higher accuracy of the model. The percentage improvement is computed in function of the lowest performing combination for each dataset.

(a) Results on the Commission for Energy Regulation (CER) of Ireland open dataset.

Normalization	Number of nearest neighbors	CRPS (CER dataset)	Percentage improvement
Per profile	100	0.231002	8.88%
Per week	100	0.238436	5.95%
Per day	100	0.239307	5.60%
No normalization	100	0.242696	4.27%
Per profile	50	0.250998	0.99%
Per profile	200	0.253509	

(b) Results on the London Low Carbon (LLC) open dataset.

Normalization	Number of nearest neighbors	CRPS (LCL dataset)	Percentage improvement
Per week	100	0.09119	4.05%
Per week	50	0.09145	3.78%
Per week	200	0.09179	3.42%
No normalization	100	0.09376	1.35%
Per profile	100	0.09427	0.81%
Per day	100	0.09504	

(c) Results on the Belgian Fluvius private dataset.

Normalization	Number of nearest neighbors	CRPS (Fluvius dataset)	Percentage improvement
Per week	200	0.030159	5.02%
Per week	100	0.030271	4.67%
Per week	50	0.030498	3.95%
Per day	100	0.030840	2.87%
No normalization	100	0.031285	1.47%
Per profile	100	0.031752	

the effects of increasing the amount of historical data on forecast accuracy.

The method deliberately excludes exogenous variables, such as weather data or household-specific attributes, as explained in Section 1.1. This decision enhances the model's generalization, its applicability to anonymized datasets, and lowers the computational complexity. This model's applicability to anonymized datasets is a crucial advantage. In many cases, smart meter data lacks specific location information. Without identifiable location data, it becomes impractical to incorporate localized weather features. This generalization allows the model to remain effective even with limited metadata, facilitating wider applicability across different datasets. Additionally, recent literature has shown that weather impact on individual household consumption patterns is less significant than on aggregated level, and does not always lead to substantial improvements in forecast accuracy. It is also important to note that when historical consumption data is used to make forecasts, the impact of weather is implicitly captured, as past consumption patterns reflect weather influences, particularly in short-term forecasts. However, excluding these exogenous data may reduce accuracy in regions or seasons where weather conditions may have a more significant impact on electricity consumption patterns. We acknowledge that weather and pricing data could play a more prominent role in household consumption forecasts with the increased penetration of solar panels, heat pumps, and dynamic pricing mechanisms. While incorporating these external variables could improve forecast accuracy, it should be noted that their inclusion would require forecasting the exogenous data itself, rather than relying on actual observed values, which introduces additional complexity. In our proposed global approach, the current input vector could be extended with the selected exogenous variables, as in [9].

Regarding the normalization step, we also consider future variations. Some approaches in the literature [58], normalize household

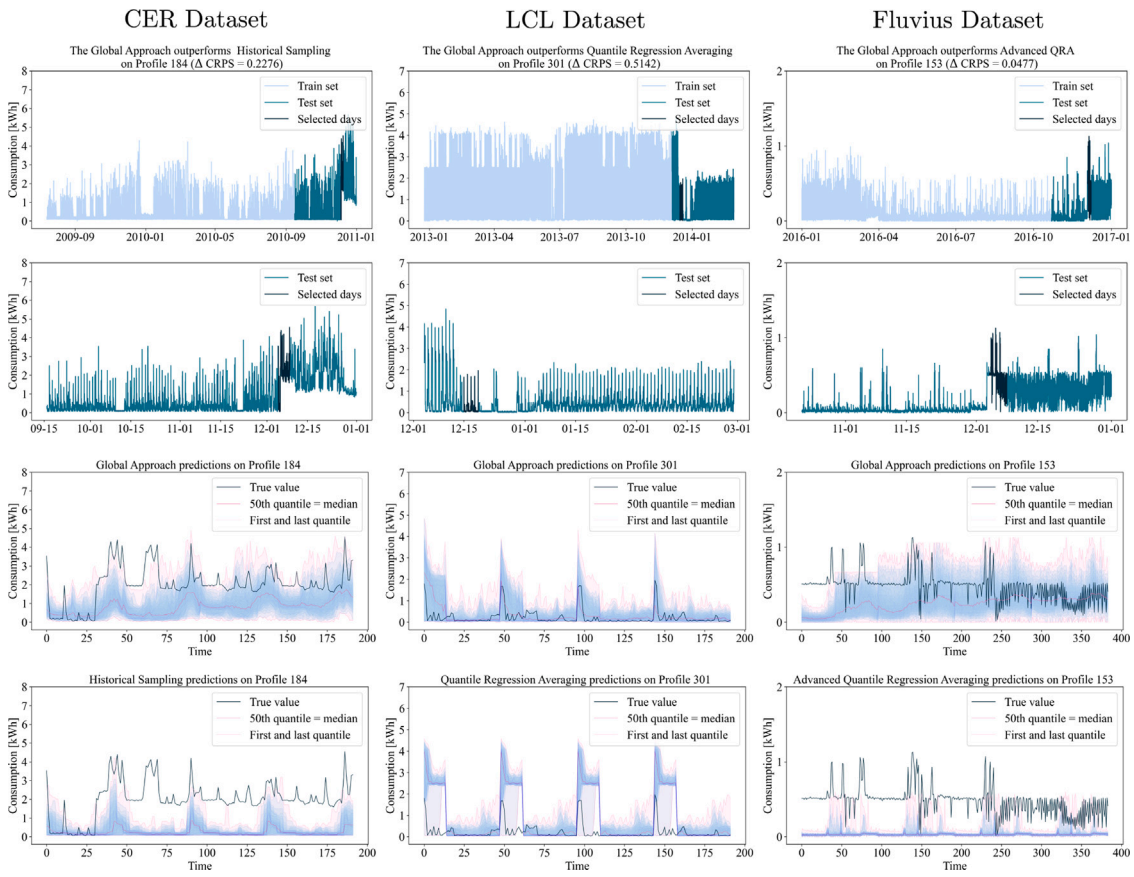


Fig. 5. The figure showcases a comprehensive analysis of three distinct household profiles, one from each dataset CER, LCL, and Fluvius. The given profiles are the ones with the largest positive difference in CRPS, indicating the instances where the global approach significantly outperforms the alternative method. The top row of subplots provides an overview of the complete profiles, where the light blue colored segments represent the training set, the dark blue colored segments represent the test set, and the black colored segment the part of the test set for which the predictions are shown in the two bottom graphs. In the second row, a closer examination of the test set alone is presented and again the black segment is the part of the test set for which the predictions are shown in the two bottom graphs. The first prediction stems from the global approach, while the second prediction reflects the alternative approach. The true values are in black. The lowest and highest quantiles are in pink, as well as the median. The intermediate quantiles are in blue. It becomes apparent that the global approach demonstrates superior adaptability to changes in consumption patterns, i.e., household profiles experiencing concept drift.

profiles using external variables like temperature. These methods adjust consumption profiles to account for weather variations. In future works, integrating exogenous data such as weather into the normalization process will be investigated. Regarding the unnormalization step, an improvement could be made by unnormalizing the consumption values based on the forecasted maximum rather than the previous day's maximum, as is currently done.

Finally, addressing concept drift actively is another avenue for future research. This could include exploring recalibrating the model as new data becomes available, online learning mechanisms or active detection schemes to improve the model's adaptability to evolving household consumption behaviors.

6. Conclusion

This paper proposes a novel probabilistic global method for forecasting residential electricity consumption based on empirical quantiles. The method does not rely on exogenous data requirements such as weather data or household-specific attributes, simplifying the implementation process. It offers the advantage of generating forecasts needing only seven days of historical data of the target household. This data requirement is relatively low compared to other approaches, making the method applicable to situations where limited historical data is available. The model's scalability is thus enhanced by its low data requirement and low complexity. Moreover, the interpretability of the model predictions is a valuable feature, as they are not derived from a black-box model, enabling stakeholders to understand and trust the forecasted results.

The proposed method demonstrates superior performance compared to state-of-the-art models on household profiles experiencing concept drift, as evidenced by the evaluation on multiple datasets. The global approach outperforms all other benchmarks on two of the datasets, the CER, and the LCL datasets. The Fluvius dataset presents unique challenges — higher skew and kurtosis, presence of negative values, shorter data span, and lower data quality, yet the proposed method still outperforms other methods on the typically challenging and changing household profiles, i.e., experiencing concept drift. This highlights the effectiveness of the method in capturing the complex dynamics of household consumption patterns.

Furthermore, this research strives for reproducibility and comparability. The utilization of several open datasets and the sharing of code contribute to making this research accessible and transparent for the scientific community. By providing these resources, researchers can replicate the experiments and build upon the findings, fostering collaboration and further advancements in the field.

Some of the design choices and advantages of the proposed method could be viewed as limitations in certain cases. For instance, while the approach only requires seven days of historical data for the target household, it also depends on a dataset containing multiple households. In cases where only a single household time series is available, the

method cannot be applied. Similarly, weather data were intentionally excluded, as recent literature shows it does not significantly enhance forecast accuracy at the household level. However, with the increasing adoption of solar panels and heat pumps, the impact of weather on individual household consumption may grow, making this an interesting avenue for future research as more relevant data becomes available.

In conclusion, this paper presents a significant contribution to the literature by introducing an accurate and global probabilistic method for short-term individual household load forecasting. The method's performance superiority, minimal data requirements, comparison with state-of-the-art approaches, and commitment to reproducibility contribute to advancing the field of probabilistic forecasting and provide a solid foundation for further research and applications in the domain of household electricity consumption forecasting.

CRediT authorship contribution statement

Lola Botman: Writing – review & editing, Writing – original draft, Visualization, Methodology, Data curation, Conceptualization. **Jesus Lago:** Supervision. **Thijs Becker:** Writing – review & editing, Validation, Supervision. **Koen Vanthournout:** Supervision, Project administration. **Bart De Moor:** Writing – review & editing, Supervision, Resources, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Lola Botman reports financial support was provided by Flemish Government (Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen). Bart De Moor reports financial support was provided by KU Leuven Research Fund. Bart De Moor reports financial support was provided by European Research Council. Lola Botman reports equipment, drugs, or supplies was provided by VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research received funding from the Flemish Government under the "Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen" program, from the KU Leuven Research Fund (projects C16/15/059, C3/19/053, C24/18/022, C3/20/117, C3I-21-00316) and from the European Research Council under the European Union's Horizon 2020 research and innovation programme (ERC Adv. Grant, grant agreement N^o 885682). The resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO), Belgium and the Flemish Government, Belgium. The work of this paper has been carried out as part of a collaboration between VITO/EnergyVille and the research group STADIUS from KU Leuven. The authors would like to thank Fluvius for sharing their data. For clarity and transparency reasons, and following the text recycling guidelines suggested in [59], the authors would like to highlight that this paper is a follow-up on [27]. Some parts of this paper have been recycled from [27] namely parts of the introduction (Section 1), of the benchmark methods BNF (Section 2.2.3), of the FFNN+Empirical Quantiles (Section 2.2.4), of the Baseline 1 (Section 2.2.5), of the performance metrics (Section 2.3) and the Irish open dataset description (Section 4.2.1).

Appendix. Statistical significance analysis of the proposed global approach vs. benchmark methods

The statistical significance of the performance differences between the global approach and the benchmark methods is assessed using the Diebold–Mariano (DM) test. The DM test is a statistical test used to evaluate whether two competing forecasts differ significantly in their accuracy, in this case based on the CRPS [47]. Note that the DM test is intended to compare forecast accuracy rather than model methods directly [60]. For each profile and for each pair of forecasting methods, the difference in CRPS values between the global approach and the alternative method is computed for each day in the test set. The mean and variance of these differences are calculated. The DM test statistic is given by:

$$\mathrm{DM} = \frac{\mathrm{mean}_{\mathrm{diff}}}{\sqrt{\frac{\mathrm{var}_{\mathrm{diff}}}{P}}}$$

where $\mathrm{mean}_{\mathrm{diff}}$ is the mean of the CRPS differences, P is the sample size (number of days in the test set), and $\mathrm{var}_{\mathrm{diff}}$ is the variance of the error differences. The p-value for the DM test statistic is computed as:

$$\text{p-value} = 2 \cdot \Pr(Z > |\mathrm{DM}|)$$

where Z represents a standard normal variable. For each method comparison, 363 p-values are computed, corresponding to the number of households in each dataset. Aggregating p-values is non-trivial due to correlations between days and between household profiles. We were not able to find a statistical test with the correct theoretical properties. Therefore, instead of interpreting an aggregated p-value, histograms of the p-values (computed per profile) are presented. Each plot in Fig. 6 compares the performances of the best methods against the second best method for each dataset. The x-axis represents the p-values, and the y-axis indicates the count of p-values within each bin. An orange vertical line marks the significance threshold of 0.05 as a reference. The percentage of p-values below this threshold is displayed in the top right corner of each plot.

In the CER and LCL datasets, the global approach generally ranks higher in average CRPS scores. The number of p-values below the threshold of 0.05 for the comparison between the global approach and the second-best alternative for the CER dataset is relatively large, suggesting a statistically significant difference for most of the profiles (Fig. 6(a)). For the LCL dataset, the same observation and conclusion can be made (Fig. 6(b)). These results aligns with the CRPS rankings. For the Fluvius dataset, the advanced Quantile Regression Averaging (QRA) method is compared against the second best benchmark. Only 28% of the profiles seem to show a statistically significant difference (Fig. 6(c)). In comparison to the global approach, the advanced QRA method shows a statistically significant difference for about half of the profiles (Fig. 6(d)).

Data availability

The code will be available in a GitLab repository upon acceptance, the link is provided in the paper. Two of the three datasets used are public and their access link is provided in the paper.

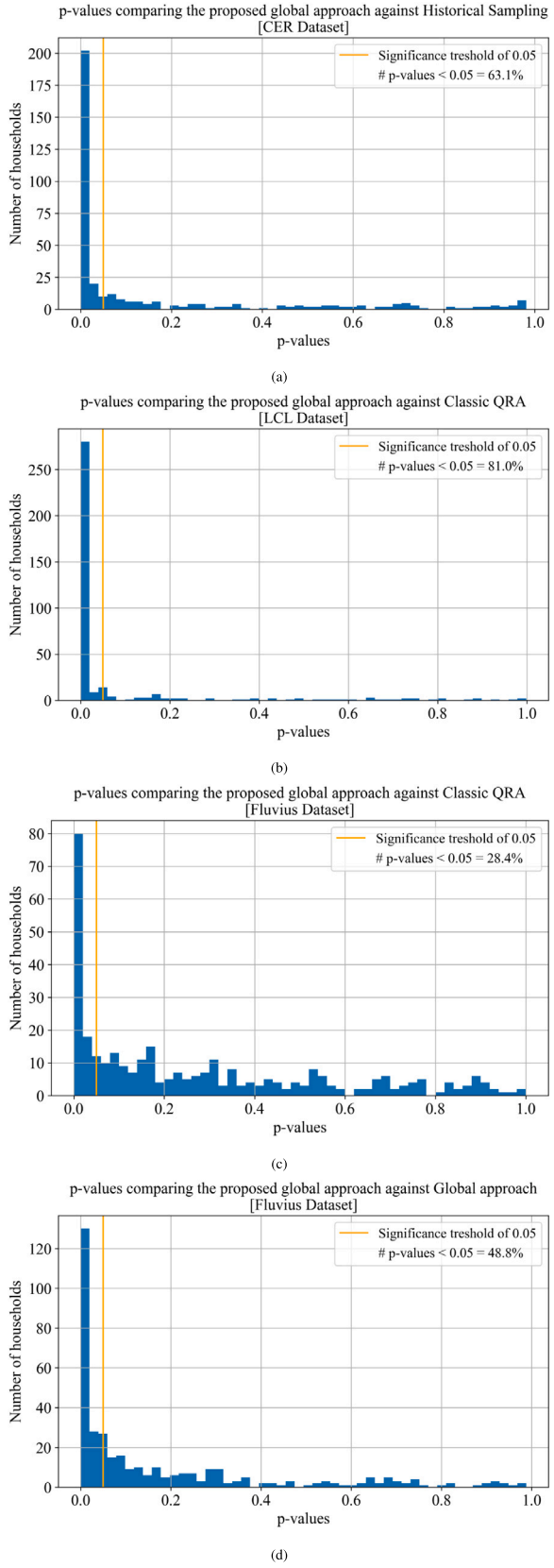


Fig. 6. Each plot compares the performances of one method against another for the three dataset. The *x*-axis represents the p-values, and the *y*-axis indicates the count of p-values within each bin. An orange vertical line marks the significance threshold of 0.05 as a reference. The percentage of p-values below this threshold is displayed in the top right corner of each plot.

References

[1] McSharry PE, Bouwman S, Bloemhof G. Probabilistic forecasts of the magnitude and timing of peak electricity demand. IEEE Trans Power Syst 2005;20(2):1166–72. http://dx.doi.org/10.1109/TPWRS.2005.846071.

[2] Haben S, Voss M, Holderbaum W. Core concepts and methods in load forecasting. 1st ed.. Springer International Publishing; 2023, p. 331. http://dx.doi.org/10.1007/978-3-031-27852-5, URL https://link.springer.com/10.1007/978-3-031-27852-5.

[3] Theodorakos K, Agudelo OM, Espinoza M, De Moor B. Decomposition-residuals neural networks: hybrid system identification applied to electricity demand forecasting. IEEE Open Access J Power Energy 2022;9(October 2021):241–53. http://dx.doi.org/10.1109/OAJPE.2022.3145520.

[4] Ben Taieb S, Hyndman RJ. A gradient boosting approach to the kaggle load forecasting competition. Int J Forecast 2014;30(2):382–94. http://dx.doi.org/10.1016/j.ijforecast.2013.07.005.

[5] Hong T, Fan S. Probabilistic electric load forecasting: A tutorial review. Int J Forecast 2016;32(3):914–38. http://dx.doi.org/10.1016/j.ijforecast.2015.11.011.

[6] Ebrahim AF, Mohammed O. Household load forecasting based on a preprocessing non-intrusive load monitoring techniques. IEEE Green Technol Conf 2018;2018-April:107–14. http://dx.doi.org/10.1109/GreenTech.2018.00028.

[7] Haben S, Ward J, Vukadinovic Greetham D, Singleton C, Grindrod P. A new error measure for forecasts of household-level, high resolution electrical energy consumption. Int J Forecast 2014;30(2):246–56. http://dx.doi.org/10.1016/j.ijforecast.2013.08.002.

[8] Haben S, Arora S, Giasemidis G, Voss M, Vukadinović Greetham D. Review of low voltage load forecasting: Methods, applications, and recommendations. Appl Energy 2021;304:1–33. http://dx.doi.org/10.1016/j.apenergy.2021.117798.

[9] Al-Qahtani FH, Crone SF. Multivariate k-nearest neighbour regression for time series data - a novel algorithm for forecasting UK electricity demand. Proc Int Joint Conf Neural Netw 2013;1–8. http://dx.doi.org/10.1109/IJCNN.2013.6706742.

[10] Dong Y, Ma X, Fu T. Electrical load forecasting: A deep learning approach based on K-nearest neighbors. Appl Soft Comput 2021;99:106900. http://dx.doi.org/10.1016/j.asoc.2020.106900.

[11] Tajmouati S, Wahbi BE, Bedoui A, Abarda A, Dakkon M. Applying k-nearest neighbors to time series forecasting: Two new approaches. J Forecast 2024;43(5):1559–74. http://dx.doi.org/10.1002/for.3093.

[12] Soenen J, Yurtman A, Becker T, D'hulst R, Vanthournout K, Meert W, Blockeel H. Scenario generation of residential electricity consumption through sampling of historical data. Sustain Energy Grids Netw 2023;34:100985. http://dx.doi.org/10.1016/j.segan.2022.100985.

[13] Jain RK, Smith KM, Culligan PJ, Taylor JE. Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. Appl Energy 2014;123:168–78. http://dx.doi.org/10.1016/j.apenergy.2014.02.057.

[14] Dang S, Peng L, Zhao J, Li J, Kong Z. A quantile regression random forest-based short-term load probabilistic forecasting method. Energies 2022;15(2):1–20. http://dx.doi.org/10.3390/en15020663.

[15] Yao R, Steemers K. A method of formulating energy load profile for domestic buildings in the UK. Energy Build 2005;37(6):663–71. http://dx.doi.org/10.1016/j.enbuild.2004.09.007.

[16] Kim TY, Cho SB. Predicting residential energy consumption using CNN-LSTM neural networks. Energy 2019;182:72–81. http://dx.doi.org/10.1016/j.energy.2019.05.230.

[17] Bessec M, Fouquau J. The non-linear link between electricity consumption and temperature in europe: A threshold panel approach. Energy Econ 2008;30(5):2705–21. http://dx.doi.org/10.1016/j.eneco.2008.02.003.

[18] Guo Z, Zhou K, Zhang X, Yang S. A deep learning model for short-term power load and probability density forecasting. Energy 2018;160:1186–200. http://dx.doi.org/10.1016/j.energy.2018.07.090.

[19] Haben S, Giasemidis G, Ziel F, Arora S. Short term load forecasting and the effect of temperature at the low voltage level. Int J Forecast 2019;35(4):1469–84. http://dx.doi.org/10.1016/j.ijforecast.2018.10.007.

[20] Elvers A, Vos M, Albayrak S. Short-term probabilistic load forecasting at low aggregation levels using convolutional neural networks. In: 2019 IEEE milan powertech, powertech 2019. 2019, http://dx.doi.org/10.1109/PTC.2019.8810811.

[21] Kang J, Reiner DM. What is the effect of weather on household electricity consumption? empirical evidence from Ireland. Energy Econ 2022;111(March):106023. http://dx.doi.org/10.1016/j.eneco.2022.106023.

[22] Xu L, Hu M, Fan C. Probabilistic electrical load forecasting for buildings using Bayesian deep neural networks. J Build Eng 2022;46(December 2021):103853. http://dx.doi.org/10.1016/j.jobe.2021.103853.

[23] Hamoudia M, Makridakis S, Spiliotis E, editors. Forecasting with artificial neural theory and applications. In: Palgrave advances in the economics of innovation and technology. 1st ed.. Springer Nature Switzerland; 2023, p. 412.

[24] Grabner M, Wang Y, Wen Q, Blažič B, Štruc V. A global modeling approach for load forecasting in distribution networks. IEEE Trans Smart Grid 2022;PP:1. http://dx.doi.org/10.1109/TSG.2023.3264525.
[25] Botman L. Python implementation: a global probabilistic approach for short-term forecasting of individual households' electricity consumption. 2022, URL https://gitlab.esat.kuleuven.be/Lola.Botman/gpsthl-forecasting/.
[26] Botman L, Soenen J, Theodorakos K, Yurtman A, Bekker J, Vanthournout K, Blockeel H, Fellow BDM, Lago J. A scalable ensemble approach to forecast the electricity consumption of households. IEEE Trans Smart Grid 2023;14(1):757–68. http://dx.doi.org/10.1109/TSG.2022.3191399.
[27] Botman L, Lago J, Becker T, Agudelo OM, Vanthournout K, De Moor B. A scalable method for probabilistic short-term forecasting of individual households consumption in low voltage grids. In: IEEE Power & Energy Society, editor. 2023 IEEE PES grid edge technologies conference & exposition (grid edge). San Diego: IEEE; 2023, p. 1–5. http://dx.doi.org/10.1109/GridEdge54130.2023.10102724, URL https://ieeexplore.ieee.org/document/10102724/.
[28] Hyndman RJ, Fan Y. Sample quantiles in statistical packages. Amer Statist 1996;50(4):361. http://dx.doi.org/10.2307/2684934, URL https://www.jstor.org/stable/2684934?origin=crossref.
[29] Ruppert D, Matteson DS. Statistics and data analysis for financial engineering. In: A computational approach to statistical learning. Springer texts in statistics, New York, NY: Springer New York; 2015, p. 75–122. http://dx.doi.org/10.1007/978-1-4939-2614-5, URL https://link.springer.com/10.1007/978-1-4939-2614-5.
[30] Kaack LH, Apt J, Morgan MG, McSharry P. Empirical prediction intervals improve energy forecasting. Proc Natl Acad Sci USA 2017;114(33):8752–7. http://dx.doi.org/10.1073/pnas.1619938114.
[31] Koenker R, Hallock KF. Quantile regression. J Econ Perspect 2001;15(4):143–56. http://dx.doi.org/10.1257/jep.15.4.143.
[32] Wang Y, Zhang N, Tan Y, Hong T, Kirschen DS, Kang C. Combining probabilistic load forecasts. IEEE Trans Smart Grid 2019;10(4):3664–74. http://dx.doi.org/10.1109/TSG.2018.2833869.
[33] Wang Y, Gan D, Sun M, Zhang N, Lu Z, Kang C. Probabilistic individual load forecasting using pinball loss guided LSTM. Appl Energy 2019;235(November 2018):10–20. http://dx.doi.org/10.1016/j.apenergy.2018.10.078.
[34] Petneházi G. Recurrent neural networks for time-series forecasting. 2019, http://dx.doi.org/10.48550/ARXIV.1901.00069, arXiv URL https://arxiv.org/abs/1901.00069.
[35] van Wyk A. Encoding cyclical features for deep learning. URL https://www.avanwyk.com/encoding-cyclical-features-for-deep-learning/%0A.
[36] Hong T, Pinson P, Fan S, Zareipour H, Troccoli A, Hyndman RJ. Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. Int J Forecast 2016;32(3):896–913. http://dx.doi.org/10.1016/j.ijforecast.2016.02.001.
[37] Liu B, Nowotarski J, Hong T, Weron R. Probabilistic load forecasting via quantile regression averaging on sister forecasts. IEEE Trans Smart Grid 2017;8(2):730–7. http://dx.doi.org/10.1109/TSG.2015.2437877.
[38] Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B Stat Methodol 1996;58(1):267–88. http://dx.doi.org/10.1111/j.2517-6161.1996.tb02080.x, URL https://onlinelibrary.wiley.com/doi/10.1111/j.2517-6161.1996.tb02080.x.
[39] Uniejewski B, Nowotarski J, Weron R. Automated variable selection and shrinkage for day-ahead electricity price forecasting. Energies 2016;9(8):621. http://dx.doi.org/10.3390/en9080621, URL http://www.mdpi.com/1996-1073/9/8/621.
[40] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. Proc ACM SIGKDD Int Conf Know Discov Data Mining 2016;13-17-Augu:785–94. http://dx.doi.org/10.1145/2939672.2939785, URL https://dl.acm.org/doi/10.1145/2939672.2939785.
[41] Barolli L. Short term load forcasting using XGBoost, vol. 1, Springer International Publishing; 2019, p. 1120–31. http://dx.doi.org/10.1007/978-3-030-15035-8.
[42] Arpogaus M, Voß M, Sick B, Nigge-Uricher M, Dürr O. Probabilistic short-term low-voltage load forecasting using bernstein-polynomial normalizing flows. In: ICML 2021 workshop on tackling climate change with machine learning. 2021, p. 1–6, URL https://www.climatechange.ai/papers/icml2021/20.
[43] Arpogaus M. Short-term probabilistic load forecasting using conditioned Bernstein-polynomial normalizing flows. 2021, URL https://github.com/MArpogaus/stplf-bnf.
[44] Arpogaus M, Voss M, Sick B, Nigge-Uricher M, Dürr O. Short-term density forecasting of low-voltage load using Bernstein-polynomial normalizing flows. 2022, p. 1–10, arXiv URL http://arxiv.org/abs/2204.13939.
[45] Shepero M, van der Meer D, Munkhammar J, Widén J. Residential probabilistic load forecasting: A method using Gaussian process designed for electric load data. Appl Energy 2018;218(March):159–72. http://dx.doi.org/10.1016/j.apenergy.2018.02.165.
[46] Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. J Amer Statist Assoc 2007;102(477):359–78. http://dx.doi.org/10.1198/016214506000001437.
[47] Gneiting T, Katzfuss M. Probabilistic forecasting. Annu Rev Stat Appl 2014;1:125–51. http://dx.doi.org/10.1146/annurev-statistics-062713-085831, URL https://doi.org/10.1146/annurev-statistics-062713-085831.
[48] Nowotarski J, Weron R. Recent advances in electricity price forecasting: A review of probabilistic forecasting. Renew Sustain Energy Rev 2018;81(June 2017):1548–68. http://dx.doi.org/10.1016/j.rser.2017.05.234.
[49] Bouyeddou B, Kadri B, Harrou F, Sun Y. DDOS-attacks detection using an efficient measurement-based statistical mechanism. Eng Sci Technol Int J 2020;23(4):870–8. http://dx.doi.org/10.1016/j.jestch.2020.05.002.
[50] Bracher J, Ray EL, Gneiting T, Reich NG. Evaluating epidemic forecasts in an interval format. PLoS Comput Biol 2021;17(2):1–15. http://dx.doi.org/10.1371/JOURNAL.PCBI.1008618.
[51] Gneiting T, Balabdaoui F, Raftery AE. Probabilistic forecasts, calibration and sharpness. J R Stat Soc Ser B Stat Methodol 2007;69(2):243–68. http://dx.doi.org/10.1111/j.1467-9868.2007.00587.x, URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2007.00587.x.
[52] Cover TM, Hart PE. Nearest neighbor pattern classification. IEEE Trans Inform Theory 1967;13(1):21–7. http://dx.doi.org/10.1109/TIT.1967.1053964.
[53] Gerossier A, Girard R, Bocquet A, Kariniotakis G. Robust day-ahead forecasting of household electricity demand and operational challenges. Energies 2018;11(12). http://dx.doi.org/10.3390/en11123503.
[54] Commission for Energy Regulation (CER). CER smart metering project - electrical customer behaviour trial. 2012, URL https://www.ucd.ie/issda/data/commissionforenergyregulationcer/.
[55] UK Power Networks. Smartmeter energy consumption data in london households. 2013, URL https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households.
[56] Rinaldi G, Garcia FC, Agudelo OM, Becker T, Vanthournout K, Mestdagh W, De Moor B. A framework for a data quality module in decision support systems: an application with smart grid time series. Int Conf Enterp Inf Syst, ICEIS - Proc 2023;1(Iceis):443–52. http://dx.doi.org/10.5220/0011749700003467.
[57] Fluvius. Honderd geanonimiseerde 15' verbruiksprofielen elektriciteit. 2021, URL https://opendata.fluvius.be/explore/dataset/1_50-aanlevering-dmk-e-kwartierwaarden/information/.
[58] Räsänen T, Voukantsis D, Niska H, Karatzas K, Kolehmainen M. Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. Appl Energy 2010;87(11):3538–45. http://dx.doi.org/10.1016/j.apenergy.2010.05.015.
[59] Hall S, Moskovitz C, Pemberton M. Text recycling: Best practices for researchers. Tech. rep. 1, 2021, p. 1–19, URL http://publicationethics.org/resources/discussion-documents.
[60] Diebold FX. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests. J Bus Econom Statist 2015;33(1). http://dx.doi.org/10.1080/07350015.2014.983236.