Applied Energy 277 (2020) 115600

Contents lists available at ScienceDirect

Applied Energy

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

Load probability density forecasting by transforming and combining quantile forecasts

Shu Zhang [a], Yi Wang [b,*], Yutian Zhang [c], Dan Wang [d], Ning Zhang [a]

[a] Department of Electrical Engineering, Tsinghua University, Beijing 100084, China
[b] Power Systems Laboratory, ETH Zurich, Zurich 8092, Switzerland
[c] China Electric Power Research Institute, Beijing 100192, China
[d] Electric Power Research Institute of Fujian Electric Power Company Limited, Fuzhou 350007, China

ARTICLE INFO

Keywords:
Probabilistic load forecasting
Ensemble learning
Quantile forecasting
Probability density forecasting
Kernel density estimation

ABSTRACT

Compared with traditional deterministic load forecasting, probabilistic load forecasting (PLF) help us understand the potential risks in the power system operation by providing more information about future uncertainties of the loads. Quantile forecasting, as a kind of non-parametric probabilistic forecasting method, has been well developed and widely used in PLF. However, the results of quantile forecasts are discrete, which contain fewer details than density forecasts which provide the most comprehensive information. This paper proposes a novel day-ahead load probability density forecasting method by transforming and combining multiple quantile forecasts. The proposed method includes two main steps: transformation and combination. In the first step, the kernel density estimation method is used to transform the individual quantile forecast into the probability density curve; in the second step, an optimization problem is established to obtain the weighted combination of different probability density forecasts. The perturbation search method is applied to determine the optimal weight of each individual forecast. We demonstrate the effectiveness and superiority of our proposed method using comprehensive case studies on the real-world load data from Guangdong province in China, ISO New England (ISO-NE) in the US and Irish smart meter data. Case studies show that the combined model is robust to kernel function selection in the transformation step and has better forecasting performance. Compared with the best individual model, the purposed combined model has an accuracy improvement of 1.54% in the Guangdong dataset and 2.9% in the ISO-NE dataset in terms of the continuous ranked probability score. The proposed combination forecasting method can be robust in high volatility scenarios.

1. Introduction

Due to the increasing integration of distributed energy resources (DER) such as rooftop photovoltaic (PV) and energy storage, the electrical load volatility increases rapidly. Traditional deterministic forecasting no longer meets our needs to accurately characterize future loads, especially for the uncertainties. Probabilistic forecasting studies the variation range of the load and the probability density in various situations. Interval, quantile, density are three main forms of probabilistic forecasting. In the face of future load fluctuations, probabilistic load forecasting (PLF) can provide more valuable information for decision making, which has become an important research direction of many scholars at present [1].

Density forecasting is the most complete form of probabilistic forecasting in the above three forms. In general, there are two ways to conduct density forecasting. One is to perform parameter estimation based on the assumed probability distribution, also known as parametric approaches; the other is non-parametric approaches [2]. For parametric estimation, a deterministic forecast is needed to provide the basis. The error of the deterministic forecast is fitted to an assumed density function. Irwin et al. used Weibull distribution to process energy billing data [3]. Herman and Kritzinger [4] fitted Gaussian, Erlang, Weibull, and Beta density function to distributions of the prediction error. Ghosh et al. compared three models based on Gaussian, log-normal and Beta distribution to generate load data in distribution system state estimation [5]. In [6], Beta distribution was further studied to discover the residential consumer load's uncertainty. To compensate for the arbitrariness of the specific distribution assumption, the Gaussian mixture model (GMM) mixes different types of distributions. It was used in the system load forecast to obtain the load probability distribution in [7]. However, because the actual load is changing and multiplicity,

* Corresponding author.
E-mail address: yiwang@eeh.ee.ethz.ch (Y. Wang).

https://doi.org/10.1016/j.apenergy.2020.115600
Received 16 April 2020; Received in revised form 9 June 2020; Accepted 24 July 2020
Available online 4 August 2020

assumed density functions are usually sub-optimal for various load conditions [8,9]. Therefore, using an assumed distribution that does not describe the true distribution well, the reliability of forecasting results cannot be guaranteed.

Getting rid of restrictive pre-assumptions on the distributions, non-parametric estimation has received a lot of attentions recently. Quantile forecasting, as a kind of non-parametric estimation tools, has been widely used in PLF. Liu et al. performed quantile regression averaging on several sister point forecasts [10]. In [11], the Gaussian process quantile regression (GPQR) model was proposed to handle load uncertainty in a non-parametric way. Hong et al. launched the 2014 Global Energy Forecast Competition (GEFCom2014) and promoted the rapid development of quantile forecasting [12]. In GEFCom2014, the team with the highest accuracy adopted the quantile generalized additive models (GAM) [13], which showed the effectiveness of the quantile forecasting method. By adding non-linear time-varying trend variables to GAM, the partially linear additive quantile regression (PLAQR) model significantly improved prediction accuracy [14]. In recent years, a large number of machine learning methods have been emerged for forecasting. Many machine learning techniques used in point forecasting can be easily applied to quantile forecasting. For decision trees and its derived models, Taieb et al. used a gradient boost regression tree method to construct a quantile forecasting model [15]. Meinshausen et al. introduced a method using random forest in quantile form [16]. The latest LightGBM also improved the support for the quantile loss function, making LightGBM quantile forecasting much easier. For neural network models, Using quantile loss function to guide the training process, a traditional LSTM network was applied to probabilistic forecasting in the form of quantiles [17]. In [18], a quantile regression neural network (QRNN) with parameter embedding was established for PLF. In [19], an improved quantile regression neural network (iQRNN) was proposed, which was more accurate, stable, and computationally efficient than traditional QRNN. In [20], a new model named LASSO-QRNN was proposed to generate quantile forecasts, and the superiority of the method was proved through the experiment on two real-world datasets.

Because quantile forecasts provide less information than density forecasts, meticulous depiction of load variation may be abandoned. By transforming quantile forecasts into density forecasts, density results can provide a holistic and detailed perspective on the load uncertainty. Kernel density estimation (KDE) is a non-parametric estimation method proposed by Rosenblatt [21] and Parsons [22]. It estimates the probability density function (PDF) without any prior knowledge or relevant assumptions about the sample distribution. Therefore, KDE is a great bridge between quantile forecast and density forecast. Haben et al. showed that the application of KDE to PLF may bring about the improvement of performance [23]. He et al. used KDE with different kernels to transform the quantiles obtained by support vector regression into density forecasts [24].

Instead of focusing on one specific model, model combination can often enhance the performance of the final model. In [25], a hybrid ensemble method was developed via integrating six primary ensemble algorithms and was proved to effectively improve the generalization ability in low voltage load forecasting. In [26], Pierro et al. showed when several data-driven forecasting methods were combined, there was significant potential in accuracy improvement. In [27], Wang et al. combined different quantile forecasting models to minimize the quantile loss. Finally, the combined model had about 4% accuracy improvement over the best individual model.

In addition to the model generation, transformation and combination of model, the evaluation of probabilistic forecasting model is very important. Different from deterministic forecasting, the evaluation of probabilistic forecasting is multi-dimensional. Reliability, sharpness, and resolution are three effective evaluation dimensions in PLF [1]. Quantile loss function, which comprehensively evaluates the reliability and sharpness of the model, is designed for quantile forecast [2]. For density forecasting models, we introduce the continuous ranked probability score (CRPS) as a metric to evaluate the sharpness and reliability of the model [28,29]. Previous studies showed that the use of CRPS could more effectively reflect the performance of the model as a whole rather than the performance at a single quantile [15,30].

Inspired by the KDE based density transformation and ensemble learning in machine learning, this paper proposes a novel day-ahead load probability density forecasting method by transforming and combining multiple quantile forecasts. Different quantile regression methods are firstly adopted to generate a series of quantile forecasts. Then, we have studied the selection of kernel function in KDE and the ensemble method of weight optimization. To determine the optimal weights, we propose the perturbation search algorithm to search weights through iterative refinement. Finally, we construct case studies based on the real-world load data from Guangdong province in China, ISO-New England in the US and smart meter data from Ireland.

The main contributions of this paper can be summarized as follows:

1. Proposing a practical method for the transformation and combination from quantile forecasts into density forecasts;
2. Designing the optimal weight determination approach, and using perturbation search for iterative solution;
3. Demonstrating the effectiveness and superiority of the proposed method by case studies with three real-world data sets.

The rest of this paper is organized as follows: Section 2 introduces the framework of our proposed method; Section 3 introduces different quantile regression models for generating individual quantile forecasts; Section 4 shows the transformation from quantile forecasts into probability density forecasts; Section 5 proposes a perturbation search algorithm to determine the optimal weights for different individual forecasts; Section 6 conducts the case studies on real-world load data from different datasets; Section 7 draws the conclusions.

2. Problem statement and framework

Transforming and combining quantile forecasts into density forecasts includes four aspects: model generation, model transformation, model combination, and model evaluation, as shown in Fig. 1.

Accordingly, there are four corresponding issues for the above four aspects:

1. Generate a series of non-parametric forecasting models. These models should contain enough information about future load uncertainties. In addition, these models need to be as independent as possible to improve the generalization of forecasts.
2. Transform the discrete quantiles into continuous density curve. The transformation process should avoid involving parameter estimation. In addition, this method should be able to give different PDF shapes for input data with different distributions.
3. Combine multiple forecasting models to a probability density forecasting model. On this issue, we need to explore a computable and tractable weight optimization algorithm.
4. Evaluate the combined model on the test datasets. The evaluation index should be able to comprehensively reflect the performance of the PLF model.

The above issues will be addressed in the following sections.

3. Quantile forecasts generation

This section briefly introduces five different quantile forecasting methods (Q-LR, Q-RF, Q-GBRT, Q-LGBM, Q-GRU) that are used for generating different individual quantile forecasts.

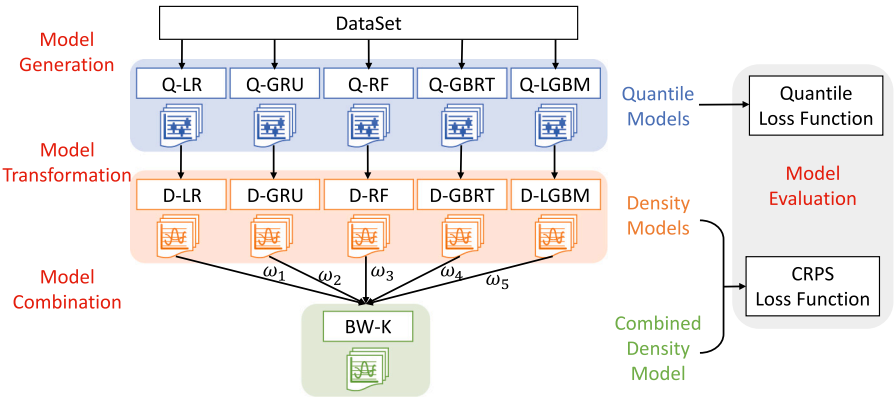

Fig. 1. Procedures of the proposed method that transforms and combines multiple quantile forecasts into a combined density forecast.

3.1. Quantile regression

Quantile regression (QR) estimates parameters for each quantile under the guidance of quantile loss that is the sum of asymmetric absolute residuals. QR model is trained by solving an optimization problem that minimizes the quantile loss:

$$\mathbf{W}_{n,q} = \arg\min_{\mathbf{W}_{n,q}} \sum_{t=1}^{T} L_{n,t,q}\left(y_t, f_{n,q}(\mathbf{X}_{n,t}, \mathbf{W}_{n,q})\right). \tag{1}$$

For each quantile q and regression model n, the quantile forecasting model $f_{n,q}\left(\mathbf{X}_{n,t}, \mathbf{W}_{n,q}\right)$ can be established. The forecasting result of the model is denoted as $\hat{y}_{n,t,q}$. In the Eq. (1), $\mathbf{W}_{n,q}$ is the parameter to be optimized; $\mathbf{X}_{n,t}$ is the input feature vector and y_t is the real load at time t; $L_{n,t,q}\left(y_t, \hat{y}_{n,t,q}\right)$ is the quantile loss of the model at time t:

$$L_{n,t,q}\left(y_t, \hat{y}_{n,t,q}\right) = \begin{cases} \left(y_t - \hat{y}_{n,t,q}\right) \times q & \hat{y}_{n,t,q} \le y_t \\ \left(\hat{y}_{n,t,q} - y_t\right) \times (1-q) & \hat{y}_{n,t,q} > y_t. \end{cases} \tag{2}$$

The quantile loss function, also known as pinball loss function, is an indicator to measure the performance of the quantile regression model. The smaller the quantile loss, the better the quantile regression model.

By varying the value of q, a total of Q quantiles of the electric load during the time T can be obtained. The average quantile loss $\overline{L}_n$ can be used to represent comprehensive performance of the n-th quantile forecasting model:

$$\overline{L}_n = \frac{1}{T \times Q} \sum_{t=1}^{T} \sum_{q=1}^{Q} L_{n,t,q}\left(y_t, \hat{y}_{n,t,q}\right). \tag{3}$$

3.2. Quantile regression models

1. **Q-LR:** Quantile linear regression (Q-LR) is the combination of quantile loss and linear regression. This method assumes the linearity between explanatory and explained variables. For Eq. (1), the regression model $f_{n,q}(\mathbf{X}_{n,t}, \mathbf{W}_{n,q})$ can be formulated as:

$$\hat{y}_{n,t,q} = f_{n,q}(\mathbf{X}_{n,t}, \mathbf{W}_{n,q}) = \mathbf{X}_{n,t} \cdot \mathbf{W}_{n,q}. \tag{4}$$

where $\mathbf{X}_{n,t} = [d_{t-H}, d_{t-H-1}, d_{t-2H+1}, d_{t-2H}, d_{t-2H-1}, d_{t-3H}]$. It contains the similar time load values in the past three days, where H denotes the total number of time points in a day.

2. **Q-GRU:** Recurrent neural network (RNN) is a type of neural network that is suitable to process time-series data. RNN not only have weight connections between layers like ANN, but also weight connections between neurons in different layers. The gated recurrent unit (GRU) is a variant of RNN proposed by Cho [31]. The newly introduced "reset" gate $\mathbf{R}_{t,q}$ and "update" gate $\mathbf{Z}_{t,q}$ control the memory and the update operation of the unit, respectively. Specifically, for each quantile q, the principle of GRU forward propagation can be described as:

$$\begin{aligned} \mathbf{R}_{t,q} &= \sigma(\mathbf{W}_{r,q} \cdot [\mathbf{H}_{t-1,q}, \mathbf{X}_{n,t}]) \\ \mathbf{Z}_{t,q} &= \sigma(\mathbf{W}_{z,q} \cdot [\mathbf{H}_{t-1,q}, \mathbf{X}_{n,t}]) \\ \widetilde{\mathbf{H}}_{t,q} &= \tanh(\mathbf{W}_{\widetilde{\mathbf{H}},q} \cdot [\mathbf{R}_{t,q} * \mathbf{H}_{t-1,q}, \mathbf{X}_{n,t}]) \\ \mathbf{H}_{t,q} &= (1 - \mathbf{Z}_{t,q}) * \mathbf{H}_{t-1,q} + \mathbf{Z}_{t,q} * \widetilde{\mathbf{H}}_{t,q} \\ y_{n,t,q} &= \sigma(\mathbf{W}_{o,q} \cdot \mathbf{H}_{t,q}), \end{aligned} \tag{5}$$

where square brackets indicate that the two vectors are connected; asterisk indicates the product of the matrix; σ and $tanh$ represent sigmoid and tanh activation functions respectively; $\widetilde{\mathbf{H}}_{t,q}$ represents new information at time t in the candidate hidden layer; $\mathbf{H}_{t,q}$ represents updated information at time t in the final hidden layer; $\mathbf{H}_{t,q}$ passes the dense layer with the sigmoid activation function, and finally outputs $y_{n,t,q}$. The input vector $\mathbf{X}_{n,t} = [h, W, M, d_{t-H}, d_{t-H-1}, d_{t-2H+1}, d_{t-2H}, d_{t-2H-1}, d_{t-3H}]$, where h is the order of the time point in a day, W is the order of the day in a week, and M is the order of the month in a year.
Constrained by the range of activation functions, normalization processing is needed to bring all input features into the range $[0, 1]$. For Q-GRU, the number of neurons in the hidden layer is the hyper-parameter that needs to be preset before the training. Through grid search, we can choose the optimal neuron quantity for different regions.

3. **Q-RF:** Random forest (RF) is an ensemble learning model based on the classification and regression tree (CART). It uses the bootstrap aggregating (Bagging) method to combine multiple weak classifiers into a strong classifier. Due to the adoption of the ensemble algorithm, it usually has higher precision than a single method, and can effectively deal with high-dimensional features without complicated hyper-parameter tuning process.
For different decision trees $T_{n_{RF}}(\cdot)$, there are different weight parameters $\mathbf{W}_{n_{RF}}$. The input feature vector $\mathbf{X}_{n,t}$ is the same as that in Q-GRU. After the decision trees are trained, each tree can output a prediction result $\hat{y}_{t,n_{RF}}$. Finally, the expectation can be obtained by averaging all the decision trees:

$$\hat{y}_{t,n_{RF}} = T(\mathbf{X}_{n,t}, \mathbf{W}_{n_{RF}}) \tag{6}$$

$$\hat{y}_t = \frac{1}{N_{RF}} \sum_{n_{RF}=1}^{N_{RF}} \hat{y}_{t,n_{RF}}. \tag{7}$$

In the above process, RF does not explicitly predict quantiles. Since RF has given multiple predictions $\hat{y}_{t,n_{RF}}$ for the same input vector, the prediction of each tree can be seen as a forecasting value. The results of each tree can be aggregated into a conditional distribution $P(Y \le y | X = x)$, thereby quantile forecasts can be obtained by empirical cumulative distribution function (CDF).

For Q-RF, there are three hyper-parameters that need to assign. With grid search, the total number of trees, minimum sample segmentation, and maximum depth are selected optimally. Since different quantiles are empirical samples, to reduce the deviation, the total amount of decision trees should not be too small. So we limit the total number of trees to no less than 500.

4. **Q-GBRT:** Gradient boosting regression tree (GBRT) is another ensemble regression model based on decision trees. Unlike RF model, each tree in the GBRT creates a new model to minimize residual of the previous training. Through iterative solution, the accuracy and generalization ability of model are gradually improved. In training, the quantile loss function $L_{n,t,q}\left(y_t, \hat{y}_{n,t,q}\right)$ is used. For the n_{GB}-th iteration, the calculation process is shown in Eq. (8):

$$f_n^{n_{GB}}(\mathbf{X}_{n,t}) = f_n^{n_{GB}-1}(\mathbf{X}_{n,t}) + \lambda \sum_{m_{GB}=1}^{M_{GB}} \xi_{m_{GB},n_{GB}} I(\mathbf{X}_{n,t} \in R_{m_{GB},n_{GB}}), \tag{8}$$

where M_{GB} is the total number of terminal nodes; $R_{m_{GB},n_{GB}}$ denotes the M_{GB} disjoint regions; $\mathbf{X}_{n,t}$ denotes the same input vector as Q-GRU; $\xi_{m_{GB},n_{GB}}$ denotes the optimal terminal node, which is calculated as:

$$\xi_{m_{GB},n_{GB}} = \arg\min_{\xi} \sum_{\mathbf{X}_{n,t} \in R_{m_{GB},n_{GB}}} L\left(y_t, f_n^{n_{GB}-1}(\mathbf{X}_{n,t}, \mathbf{W}_{n,q,n_{GB}-1}) + \xi\right). \tag{9}$$

Q-GBRT involves a complex hyper-parameter adjustment process. The number of iterations (i.e. the number of trees) and the maximum depth are the main hyper-parameters. Similarly, grid search method used for each region to optimize the hyper-parameter selection. In addition, to reduce over-fitting, the minimum number of leaves per node and splits, the maximum number of features and the sub-sample ratio also need to be carefully chosen.

5. **Q-LGBM:** LightGBM is a distributed improvement of traditional GBRT. Thanks to the histogram method, LightGBM can significantly reduce computational cost while facilitating parallel operations. Due to the special way the data is processed, results of Q-LGBM differ from that of GBRT. Hence, LightGBM will also be used as an independent method to participate in our research. For Q-LGBM, the accuracy mainly depends on the number of leaves n_{leaf}. By adjusting the parameters to limit the maximum depth, selecting an appropriate sample sampling ratio and minimum split weight, over-fitting can be greatly suppressed.

4. Kernel density estimation based forecast transformation

Based on the quantile forecasting results, we can transform a series of quantiles into a continuous density curve. KDE method estimates the unknown density function from data itself. It uses a kernel function to fit the data points to form an optimal estimate of the true probability distribution. The KDE result can be expressed as:

$$\hat{k}_{n,t}(x) = \frac{1}{Qw} \sum_{q=1}^{Q} K\left(\frac{x - \hat{y}_{n,t,q}}{w}\right), \tag{10}$$

where w is bandwidth; $\hat{y}_{n,t,q}$ is the quantile forecasting result; $K(\cdot)$ is the kernel function.

In theory, when there are enough points input, the loss of efficiency is small for the different kernels [32]. However, when the number of input points is limited, the choice of the kernel functions matters. In previous studies, Gaussian kernel [33], Epanechnikov kernel [34], uniform kernel [20] and triangular kernel [35] have been widely used in KDE. Table 1 compares the form and domain of these kernel functions.

Table 1
The function form and domain of four common used kernel functions.

Type	Gaussian	Epanechnikov	Triangular	Uniform
Form	$\frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$	$\frac{3}{4}(1-x^2)$	$(1-\lvert x\rvert)$	$\frac{1}{2}$
Domain	$(-\infty, +\infty)$	$[-1, +1]$	$[-1, +1]$	$[-1, +1]$

Since system load cannot be negative, the predicted probability distribution range should not contain the negative part. If we simple select positive part and re-normalization, the estimation usually underestimates the density. This is because estimator cannot feel the boundary, and penalizes for the lack of data on the negative axis. The most direct idea is that we put some mirror data on the negative axis. This method of boundary correction is called the reflection method. The reflection method adds reflection data centered on the boundary, and then we estimate a new distribution. Eq. (11) is the KDE result for $x \geq 0$. For $x < 0$, $\hat{k}_{n,t}(x)$ will become 0.

$$\hat{k}_{n,t}(x) = \frac{1}{Qw} \sum_{q=1}^{Q} \left[K\left(\frac{x + \hat{y}_{n,t,q}}{w}\right) + K\left(\frac{x - \hat{y}_{n,t,q}}{w}\right)\right]. \tag{11}$$

After carrying out the model transformation, five density forecasting models (D-LR, D-GRU, D-RF, D-GBRT, D-LGBM) can be obtained. All models give predictions on the same load value, which benefits the following combination work.

5. Optimal weighted combination

Model combination can improve the generalization ability of the model, and may improve the accuracy and robustness of the model. This section discusses three issues about model combination: density forecast performance metrics, combination method, and program algorithm to achieve optimal weights. At the same time, five competing methods are introduced as the benchmark.

5.1. Metrics: Continuous Ranked Probability Score (CRPS)

CRPS is a comprehensive index which evaluates the reliability and sharpness of density forecasting model [28]. Fig. 2 shows the calculation of the CRPS function. The CPRS does not focus on any particular point of the probability distribution but considers the predicted distribution as a whole. CRPS evaluates the square of the differential area between the CDF of observation and the CDF of prediction. The smaller the score, the closer the probability distribution is to the true distribution, and the better the performance of the forecast. CRPS generalize the mean absolute error (MAE): while the forecasts are probabilistic, the observations are deterministic. If the forecast is deterministic, it is reduced to the MAE. For density forecasting model n at time t, X is a random variable of the predicted probability distribution, $F_{n,t}(y) = \mathbf{P}[X \leq y] = \int_{-\infty}^{y} \hat{k}_{n,t}(\lambda) d\lambda$ represents the CDF of the random variable X, where $\hat{k}_{n,t}(\lambda)$ is the PDF of X. The CRPS for model n at time t is defined as

$$CRPS_{n,t}(F_{n,t}(y), y_t) = \int_{-\infty}^{+\infty} (F_{n,t}(y) - \mathbf{I}(y - y_t))^2 dy, \tag{12}$$

where y_t is the true value; $\mathbf{I}(y - y_t)$ is the Heaviside step function. If $y \geq y_t$, $\mathbf{I}(y - y_t)$ is 1, otherwise 0.

5.2. Method: Probabilistic load forecasting model combination

Model combination is a comprehensive consideration of the forecasting results of each model. We weighted the model as a whole by choosing weights for each model. Based on the individual density models, the PDF of the combined density model $\hat{k}_t(x)$ can be calculated as

$$\hat{k}_t(x) = \sum_n \omega_n \hat{k}_{n,t}(x). \tag{13}$$

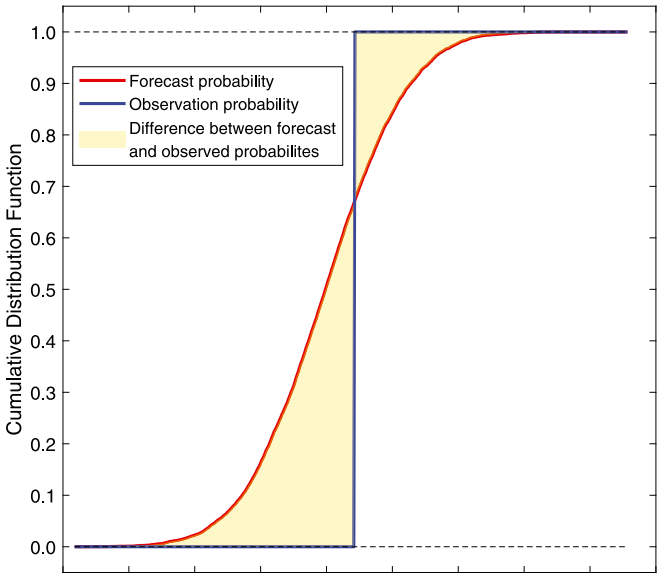

Fig. 2. Schematic diagram of CRPS function.

To find the optimal weight, we convert the weight determination procedure into an optimization problem to minimize CRPS:

$$\begin{aligned}\omega_n &= \arg\min_{\omega_n} \frac{\sum_{t=1}^{T} CRPS_t(F_t(y), y_t)}{T} \\ &= \arg\min_{\omega_n} \frac{\sum_{t=1}^{T} CRPS_t(\int_{-\infty}^{x} \sum_n \omega_n \hat{k}_{n,t}(\lambda) d\lambda, y_t)}{T} \\ \text{s.t.} \quad &\sum_{n=1}^{N} \omega_n = 1, \omega_n \geq 0 \quad \forall n \in [1, N],\end{aligned} \tag{14}$$

where $CRPS_t(F_t(y), y_t)$ is the CRPS value under $F_t(y)$ of the combined model; $\hat{k}_{n,t}(\cdot)$ is the KDE process introduced in Section 4.

As shown in Eq. (12), there is no explicit analytical expression for the objective function to minimize CRPS. Therefore, common optimization methods, such as gradient method and quasi-Newtonian method, are hard to be applied to this problem. Therefore, the difficulty of the combination lies in how to achieve optimal weight efficiently.

5.3. Algorithm: Perturbation search

The main idea of perturbation search is to change the weight of each model slightly, find the maximum improvement direction, and keep iterating to improve the performance until no improvement can be made. Fig. 3 is the diagram of the iteration process. To be specific, one round of perturbation search consists of three steps: (1) start from initial weights obtained in the last round; (2) add small disturbances to different weights and normalize the weights to form a set of candidates; (3) find the weights of the model with the smallest CRPS as the initial weight for the next iteration.

Using perturbation search algorithm, the detail calculation process of the purposed method is provided in Algorithm 1:

Algorithm 1: Perturbation Search Based Model Combination Method

Input: initial weight $\mathbf{W}^{(0)} = \left[\omega_1^{(0)}; \omega_2^{(0)}, ..., \omega_N^{(0)}\right]$; step length (disturbance) ϵ; density forecasting result $\hat{k}_{n,t}(x)$; real value y_t
Output: optimal weight $\mathbf{W}^* = [\omega_1^*, \omega_2^*, ..., \omega_N^*]$
Initial:
Define iteration count $i = 0$; temp vector
$\mathbf{W}_{temp} = [w_1, w_2, ..., w_N] = \mathbf{W}^{(0)}$; Calculate $CRPS^{(0)}$
repeat
  Increment the counter value and update the weights:
  i=i+1
  $\mathbf{W}^{(i)} = \mathbf{W}_{temp}$
  for *n=1 to N* **do**
    Load previous weights:
    $\mathbf{W}_{temp} = \mathbf{W}^{(i)}$
    Add small perturbations to the model n:
    $w_n = w_n + \epsilon$
    Normalize so that the sum of the weights is 1:
    $\mathbf{W}_{temp} = \frac{\mathbf{W}_{temp}}{\sum_{n=1}^{N} w_n}$
    for *t=1 to T* **do**
      Obtain the combined density model's PDF:
      for *All x* **do**
        $\hat{k}_t(x) = \sum_{n=1}^{N} w_n \hat{k}_{n,t}(x)$
      end
      Integrate to get CDF:
      $F_{n,t}(x) = \int_{-\infty}^{x} \hat{k}_t(\lambda) d\lambda$
      Calculate $CRPS_{n,t}^{(i)}$ based on the result of CDF:
      $CRPS_{n,t}^{(i)} = \int_{-\infty}^{+\infty} (F_{n,t}(x) - \mathbf{I}(x - y_t))^2 dx$
    end
    Calculate the average CRPS of the model with increasing weight of model n:
    $CRPS_n^{(i)} = \frac{\sum_{t=1}^{T} CRPS_{n,t}^{(i)}}{T}$
  end
  Label model with minimum $CRPS_n$ as $CRPS^{(i)}$ and its weights as $\mathbf{W}_{temp}$.
until $CRPS^{(i)} > CRPS^{(i-1)}$;
Output result: $\mathbf{W}^* = \mathbf{W}^{(i-1)}$

5.4. Competing methods

The order of model transformation and model combination can be reversed, and the combination strategies can vary. In this subsection, we will introduce five competing methods in comparison with our purposed method.

Because model combination can be done directly after quantile forecast generation or after KDE process, methods can be divided into two categories:

1. **Transformation first model (name ends with K):** Firstly, KDE is performed on individual quantile forecasting models to generate a total of N density forecasting models. Secondly, the PDF is weighted to obtain a combined density forecast. As shown in the Eq. (15), $\hat{k}_{n,t}(x)$ is the model n's PDF at time t, and $\hat{k}_t(x)$ is the combined density forecasting result at time t:

$$\hat{k}_t(x) = \sum_n \omega_n \hat{k}_{n,t}(x) \tag{15}$$

2. **Combination first model (name ends with E):** For each individual quantile model, a weight of ω_n is chosen firstly, and each individual model is weighted to form a new combined quantile forecasting model as Eq. (16). Then KDE is performed to get a combined density forecasting model. In this case, we do the combining work with quantiles, $\hat{y}_{n,t,q}$ is the model n's q-quantile forecasting value at time t, and $\hat{y}_{t,q}$ is the combined q-quantile value at time t.

$$\hat{y}_{t,q} \approx \sum_n \omega_n \hat{y}_{n,t,q} \qquad \forall q \in [1, 2, \ldots, Q] \tag{16}$$

Noting that weighting results at quantiles is not necessarily the real result of combined model at these quantiles, this method may have an approximation [27].

In addition to the order of transformation and combination, for each kind of order, there are three ways to determine the weight.

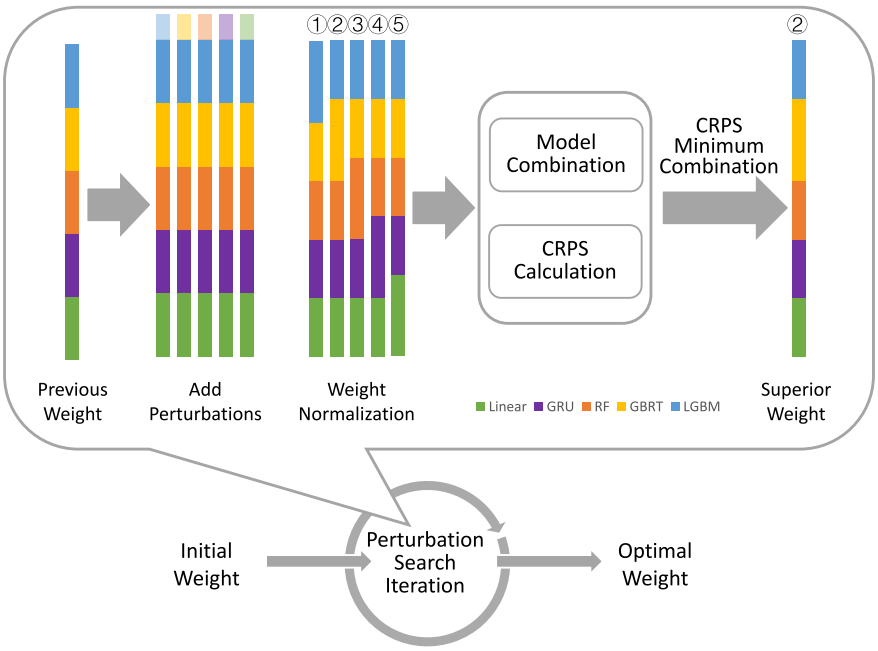

Fig. 3. An iteration perturbation search diagram.

1. Best weight (BW): The weights are determined by the CRPS-guided perturbation search algorithm as Algorithm 1.
2. Simple average (SA): This strategy applied equal weight to each model:

$$\omega_n = 1/N. \tag{17}$$

3. Weighted average (WA): Models with higher prediction accuracy have higher weights:

$$\omega_n = \frac{\frac{1}{\bar{L}_n}}{\sum_{n=1}^{N}\frac{1}{\bar{L}_n}} \tag{18}$$

According to different order and weight determination strategy, there are several competing methods:

1. **BW-E:** Compared with the purposed method, perturbation search algorithm is also used, but the combination work is prior to the transformation. A similar form of optimization problem is established. KDE $k_t(x)$ inputs the combined quantile model results and is calculated in every iteration, which may cause heavy calculation burden.

$$\begin{aligned}\omega_n &= \arg\min_{\omega_n}\frac{\sum_{t=1}^{T} CRPS_t(F_t(y), y_t)}{T}\\ &= \arg\min_{\omega_n}\frac{\sum_{t=1}^{T} CRPS_t(\int_{-\infty}^{x} k_t(\lambda)d\lambda, y_t)}{T}\\ &= \arg\min_{\omega_n}\frac{\sum_{t=1}^{T} CRPS_t(\int_{-\infty}^{x} k_t(\lambda)\big|_{\hat{y}_{t,q}=\sum_n \omega_n \hat{y}_{n,t,q}} d\lambda, y_t)}{T}\\ \text{s.t.}\quad &\sum_{n=1}^{N}\omega_n = 1, \omega_n \geq 0 \quad \forall n \in [1, N].\end{aligned} \tag{19}$$

2. **SA-K:** Each model has equal weight, and model transformation is before model combination. The combined model is obtained by Eq. (15).
3. **SA-E:** Each model has equal weight. Combination process is like Eq. (16). After that, the single combined quantile model is transformed into the density model using KDE.
4. **WA-K:** Weighted by the reciprocal of the value of the quantile loss function and the rest of the process is similar to SA-K.

Table 2
BW-K method and other competing methods.

Priority	Strategy		
	Best weight	Simple average	Weighted average
Transformation first	BW-K(Proposed)	2.SA-K	4.WA-K
Combination first	1.BW-E	3.SA-E	5.WA-E

5. **WA-E:** Weighted by the reciprocal of the value of the quantile loss function and the rest of the process is similar to SA-E.

The method proposed in this paper can be called BW-K because perturbation search algorithm is used and transformation precedes combination. Table 2 summarizes the competing combination methods from two aspects: (1) the order of model combination and model transformation; (2) the weight determination strategies.

6. Case studies

In this section, we carry out two detailed case studies on the system load data from Guangdong Province in China and ISO-NE in the United States. Case studies of residential smart meters in Ireland are also implemented to verify the adaptability of the proposed method to high-volatility scenarios. The effectiveness and superiority of the proposed method have been fully demonstrated. The training and analysis work of this study is done on a laptop (CPU: Intel Core 6700HQ, memory: 16 GB, GPU: Nvidia GeForce 960M). The quantile forecasting model is based on the Python 3.7.2 environment, using the TensorFlow [36] 1.14.1 GPU version as the framework. The weight determination of the combined model and the assessment of the model are done in MATLAB R2019a.

6.1. Experiment setups

Using the same data for model generation and model combination has a high risk of over-fitting. To avoid this, we divide the original data set into three parts in chronological order, namely D1, D2, D3. D1 is the data set for individual model training and hyper-parameter tuning. In particular, for GRU, GBRT, RF and LGBM models, the 5-fold cross-validation method is used in D1. D2 is used to determine the weight

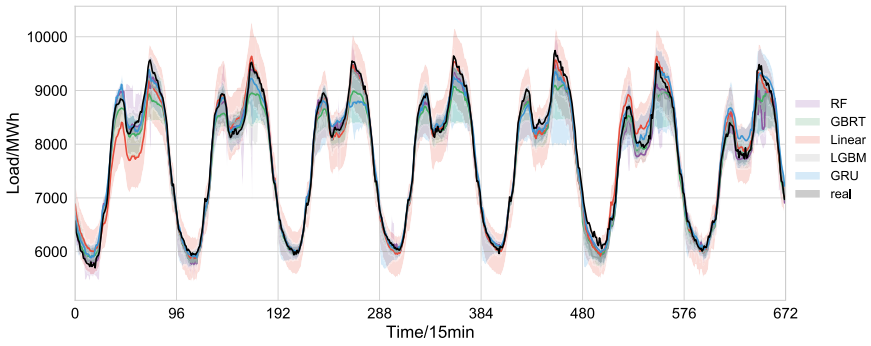

Fig. 4. Time series graph of load from Guangdong province. The solid color line is the 50-quantile forecasting value, the color shading is the range of variation between the 10-quantile and 90-quantile forecasting values, and the black line is the real load.

Table 3
Average CRPS of load from Guangdong province for each individual model and combined model. Gaussian kernel function is used for KDE.

Data sets	Models										
	RF	GRU	GBRT	LR	LGBM	SA-E	WA-E	BW-E	SA-K	WA-K	BW-K
Train Set	128.24	127.28	118.88	137.11	114.17	115.28	115.01	113.59	114.39	114.22	**113.50**
Test Set	118.53	114.25	122.92	129.33	104.75	103.92	103.85	103.14	103.17	103.03	**102.88**

of the combined model. D3 is used to evaluate the combined model. The ratio of the length of the D1: D2: D3 is simply chosen to be 2: 1: 1 for the 4-year time scale, thereby reducing the impact of monthly fluctuations of the load on the overall forecasting performance. In this study, if the load characteristics change in one of D1, D2, and D3, the result may be affected. The time span of the available residential smart meter data is less than 4 years. We divide the data set according to 2:1:1 ratio, ignoring the possible impact of seasonality. Although data split is not the scope of this article, it can be solved by building a more refined data partition model [37] and using cross-validation to divide the optimal data set [38,39].

Input feature selection is a tricky work, which impacts efficiency and accuracy. In our study, we followed two general methodologies: (1) similar day method and (2) recency effect. Lagged variables contain the similar time load values in the past three days. Category variables help models distinguish between weekdays and weekends, seasonal characteristics, and day-night characteristics. In our early study, we have tested lots of input feature portfolio. The input features with three-day lagged variables can achieve high accuracy with reasonable amount of computation.

6.2. Case studies on load data from Guangdong province

This case study is conducted on the system load from Guangdong province in China from 2003 to 2006. This data set is sampled every 15 min, which is 96 points a day. In the process of establishing quantile model, LR, GRU, RF, GBRT and LGBM model takes 1.81 s, 1002 s, 289.5 s, 983.6 s and 4.77 s. KDE takes 38.44 s, and the optimal weight selection takes 233.55 s. The optimal weight is reached after 135 iterations. The total time of this method is acceptable for the day-ahead load forecasting of a single area.

Fig. 4 provides the 7-day forecasting results for each individual model from November 27, 2006, to December 3, 2006. It can be seen that the real load is within the fluctuation range of each individual model most of the time, and the trend of the predicted value is similar to the real value. Among them, the linear model has a large prediction interval. The tree-based models have small fluctuation, but there may be a glitch in the prediction. The GRU model is usually stable, but it is not accurate enough to describe the peak load.

Fig. 5 is a PDF graph for every hour of the individual model and BW-K model on November 27, 2006. The vertical red line is the true load value, and the curves of other colors are the probability density distribution predicted by different models. It shows that the individual model closest to the true value is not the same at different times. This confirms that no individual forecasting method is optimal for all cases. From the graph, the real load mostly appears near the highest density point of the combined model's PDF, which indicates that the combined model (BW-K) is usually more stable than the single model, and reduces the overall risk of making a poor model selection.

Table 4
Average CRPS of load from Guangdong province with different kernel functions for each individual model and combined model.

Kernels	Models							
	RF	GRU	GBRT	LR	LGBM	SA-E	WA-E	BW-K
Gaussian	**118.53**	**114.25**	**122.92**	**129.33**	**104.75**	**103.92**	**103.85**	**102.88**
Epanechnikov	118.83	114.49	123.22	129.74	104.93	104.30	104.22	103.14
Triangle	118.72	114.39	123.13	129.58	104.87	104.17	104.10	103.05
Uniform	119.05	114.68	123.45	129.98	105.08	104.54	104.47	103.47

Table 5
BW-K model weight table of load from Guangdong province for combined models with different kernel functions.

Kernels	Models				
	RF	GRU	GBRT	LR	LGBM
Gaussian	0.1459	0.0527	0.1277	0.0854	0.5883
Epanechnikov	0.1462	0.0565	0.1263	0.0818	0.5892
Triangle	0.1461	0.0565	0.1269	0.0816	0.5888
Uniform	0.1480	0.0468	0.1306	0.0853	0.5893

Table 3 compares the average CRPS for each individual model and combined model. In test set, the simple average (SA-E, SA-K) and the weighted average (WA-E, WA-K) models can reduce the prediction bias, and obtain a more accurate result than the best individual model. The BW-K model has the best performance among them, and the accuracy is improved by 1.54% compared with the best individual model. Compared with the SA-E model, the BW-K is also improved by 0.75%.

Table 4 shows that both individual and combined model using KDE with Gaussian kernel has the best prediction accuracy, but the difference of precision between different kernel is very small. It shows that the combined model has strong robustness in the selection of kernel function.

Comparing the value in Table 5, we can see that under different kernel functions, the weight changes little. The BW-K model can greatly

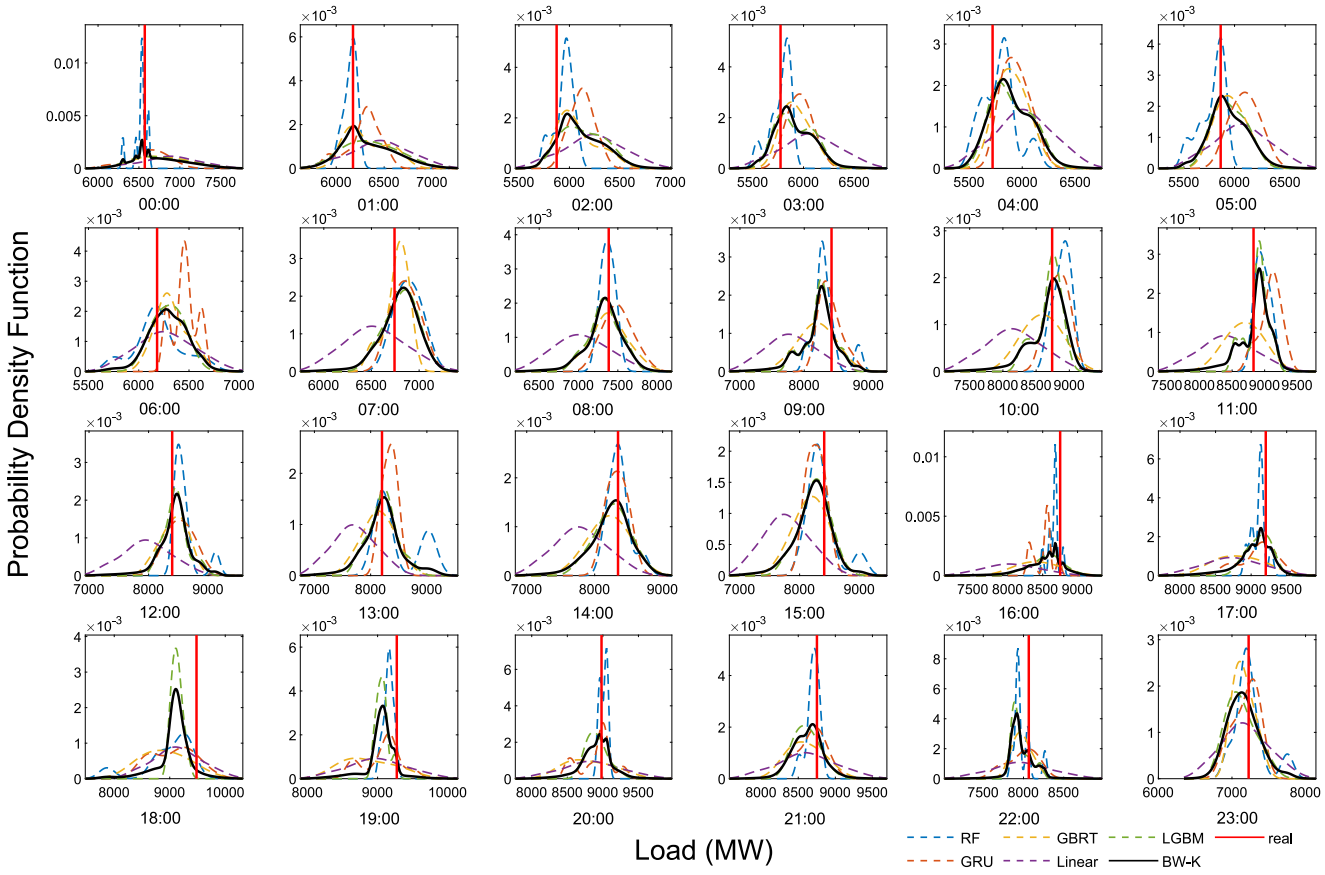

Fig. 5. The PDF graph of load from Guangdong province for each individual model and BW-K model on November 27, 2006.

adjust the weights of different models. For example, quantile linear regression (LR) has a lower weight in the model and a model with better prediction performance, such as LightGBM, has a higher weight.

6.3. Case studies on load data from ISO-NE

Another case study is conducted on the hourly load data set from ISO-NE in the US from 2013 to 2016. The data set includes eight zones: Connecticut (CT), Maine (ME), Massachusetts (SEMASS, WCMASS, NEMASS), New Hampshire (NH), Rhode Island (RI), and Vermont (VT). For different regions, the training time varies greatly, which is mainly affected by the selection of hyperparameters. On average, it takes 3506 s to build the quantile forecasting model, 59.6 s for the KDE process, and 2.63 s for each iteration. The optimal weight can be obtained in no more than 300 iterations, and the median number of iterations is 198. In general, most of the time is spent on the establishment of quantile forecasting model, while the selection of optimal weight takes less time. In summary, the time taken to run this method on a laptop is acceptable.

The process of weight optimization is the process of CRPS reduction. In order to express clearly, we use CRPS decreasing with iteration number to express the convergence of the algorithm. Fig. 6 shows that as the number of iterations increases, the CRPS of both the training set and the test set decreases and gradually converges in all eight regions. Different regions need different numbers of iterations for convergence. The initial weights of all models are equal. The greater the difference between optimal weight and equal weight, the more iterations are required. Overall, in all training sets, CRPS drops very smoothly and the model converges. In most test sets, CRPS can drop and converge very smoothly. Some test sets have a small amount of jitter or increase when the number of iterations is high. The results show that the over-fitting caused by weight selection algorithm is not serious, and the CRPS of test sets in all regions decrease to the minimum or very close to the minimum.

Table 6 illustrates the average CRPS of ISO-NE zonal-level load for each individual model and combined model. The proposed BW-K model get the best forecasting performance in all zones. In Fig. 7, the relative improvements of different combined models are showed compared with the best individual model.

On average, BW-K model has a 2.9% accuracy improvement over the best individual model, and about 1.4% over SA-E model. The improvement rate of BW-K model is up to 6.24% in NH and at least 2.05% in SEMASS. As the bar chart shows, not all model combination benefits. Both the simple average method and the weighted average method may be less effective than the best individual model. In addition, we can see that the model transformation first approach (whose name ends in K) is more accurate than the model combination first approach (whose name ends in E). In general, the best weight combined models are better than the weighted average models, and the weighted average models are better than the simple average models. Hence, more considerations on weight determination strategy can greatly improve performance.

To verify the robustness of the combined model, we try different kernel functions in the model transformation. Table 7 indicates that different kernels may slightly impact the accuracy of forecasting results. Nevertheless, the differences between kernel selection are quite small. Except for VT zone, the largest relative gap in other zones is less than 0.1%. Hence, the choice of kernel function is sufficiently robust to the transformation and combination methods proposed in this paper.

6.4. Case studies on residential smart meter data from Ireland

In order to verify that our proposed method still works in volatile time series, we conduct case study on the residential smart meter data

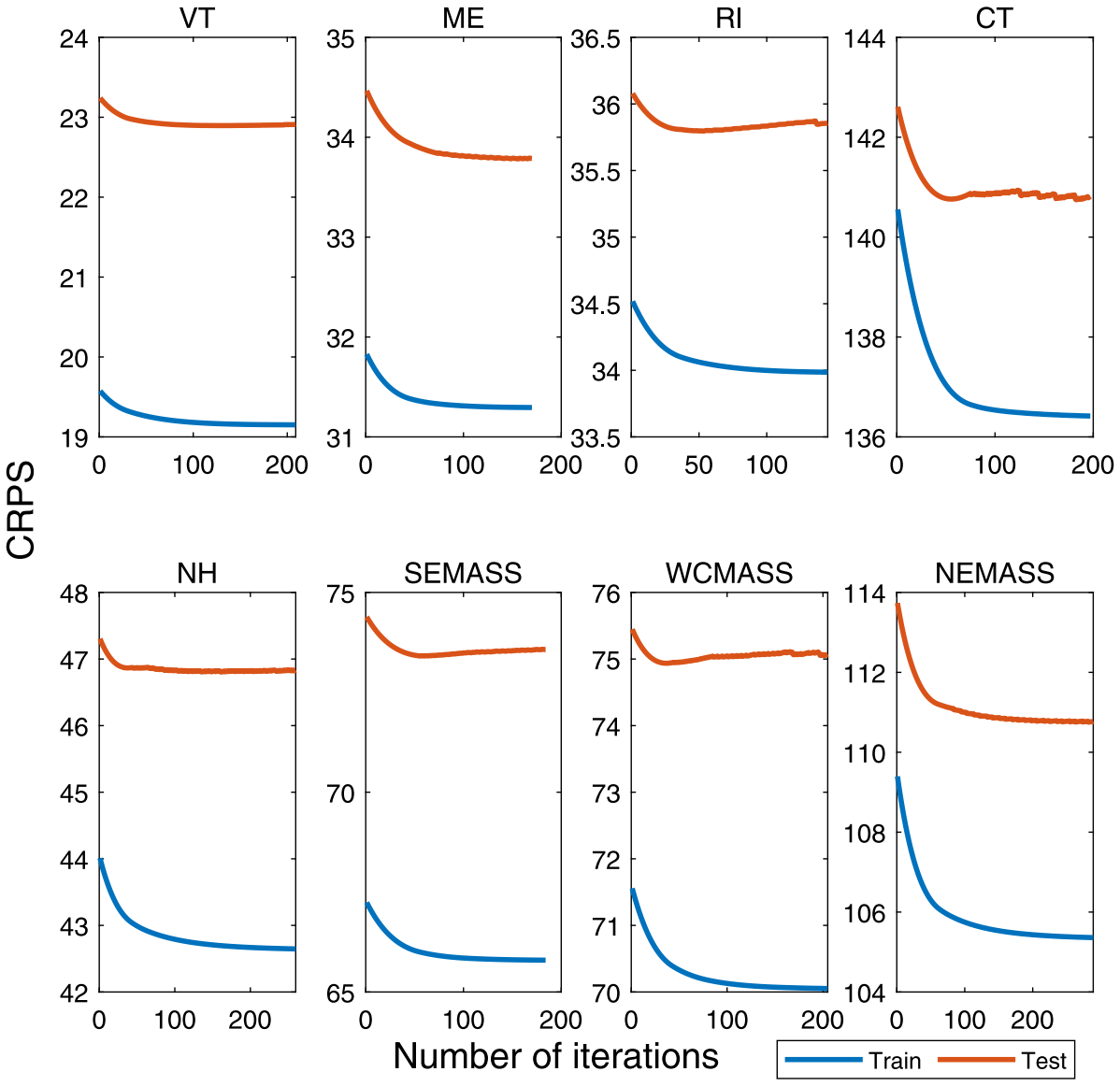

Fig. 6. CRPS for each iteration from ISO-NE. The blue line represents the CRPS result of the training set, and the orange line represents the result of the test set.

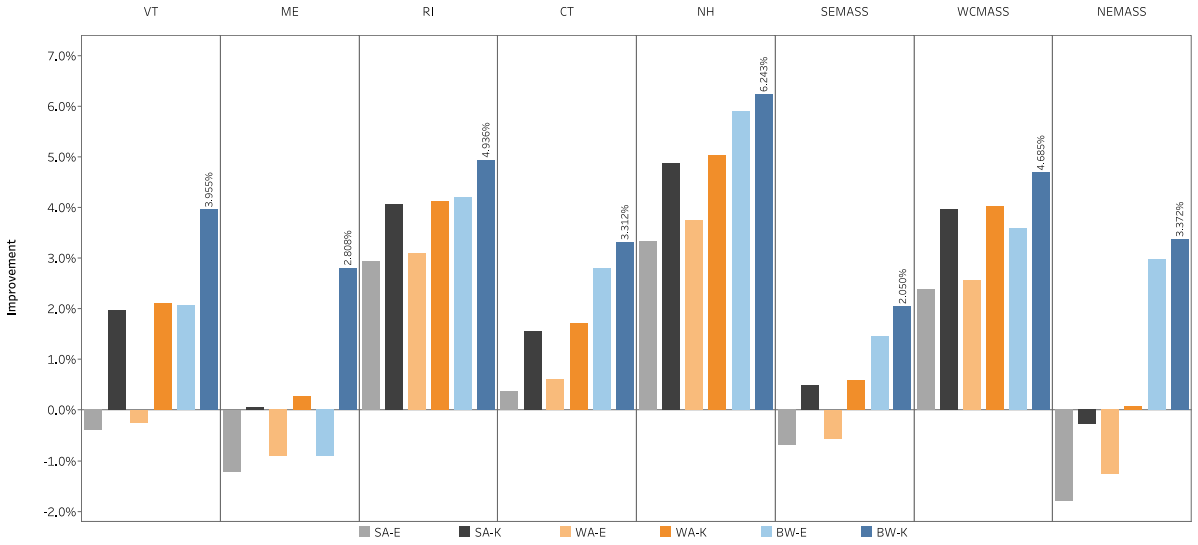

Fig. 7. The relative improvement of the different combination methods of zonal-level load from ISO-NE compared with the best individual model.

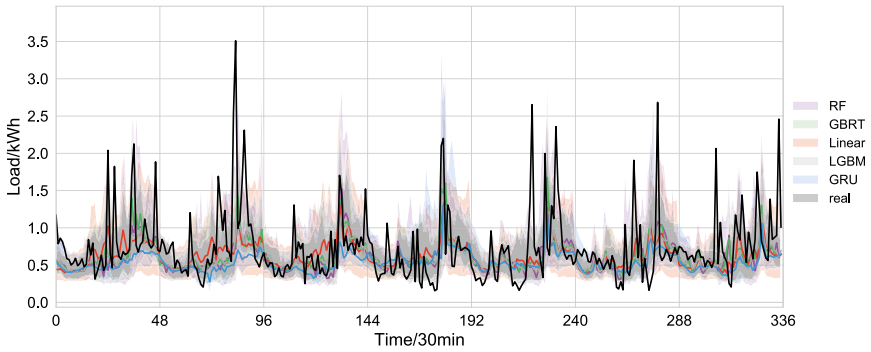

Fig. 8. Quantile load forecasts of household #1003 of one week from April 9, 2010 to April 15, 2010. The solid color line is the 50-quantile forecasting value, the color shading is the range of variation between the 10-quantile and 90-quantile forecasting values, and the black line is the real load.

Table 6
Average CRPS of zonal-level load from ISO-NE for each individual model and combined model in the test set. The Gaussian kernel function is used for KDE.

Zones	Models										
	RF	GRU	GBRT	LR	LGBM	SA-E	WA-E	BW-E	SA-K	WA-K	BW-K
VT	26.200	23.600	27.357	26.398	23.650	23.669	23.644	23.240	23.255	23.231	**22.909**
ME	39.135	34.504	39.609	40.330	34.930	34.815	34.736	34.736	34.489	34.435	**33.787**
RI	40.986	37.215	43.784	42.009	37.332	36.403	36.363	36.053	36.097	36.081	**35.855**
CT	163.653	144.354	172.922	163.352	148.512	143.949	143.698	141.371	142.699	142.522	**140.816**
NH	53.253	49.143	55.592	60.202	49.100	47.892	47.739	46.952	47.334	47.271	**46.831**
SEMASS	81.910	74.703	86.671	82.286	78.226	75.090	75.011	73.904	74.433	74.381	**73.570**
WCMASS	83.705	77.769	87.700	90.761	77.895	76.398	76.302	75.694	75.483	75.456	**75.073**
NEMASS	129.611	113.606	141.393	137.880	117.499	115.122	114.664	111.105	113.837	113.552	**110.771**

Table 7
Average CRPS of zonal-level load from ISO-NE for BW-K model, in which different kernels are used in the model transformation.

Zones	Kernels			
	Gaussian	Uniform	Triangle	Epanechnikov
VT	**22.909**	23.085	23.089	23.087
ME	33.787	**33.775**	33.779	33.776
RI	**35.855**	35.862	35.860	35.863
CT	140.816	140.734	140.761	**140.699**
NH	46.831	46.828	46.824	**46.821**
SEMASS	73.570	**73.541**	73.557	73.546
WCMASS	75.073	**75.040**	75.077	75.046
NEMASS	110.771	**110.708**	110.734	110.730

set in Ireland from July 14, 2009 to December 31, 2010. This data set is sampled every 30 min, which is 48 points a day. Five household smart meter data sets are selected for the study. For the five households, the mean time of quantile forecasting model establishment, quantile model transformation to probability density model, and weight selection iteration per round are 607 s, 44.8 s and 1.89 s, respectively. The optimal weight can be achieved within 350 iterations in all households, with a median of 196. Like the above two case studies, the time cost of running the model is completely acceptable.

Fig. 8 shows the prediction results of each quantile model for #1003 household smart meter data. Different from the forecast of system load, the fluctuation of resident data increases significantly, and the real value will break the forecast upper and lower limits of 90-quantile and 10-quantile in many places.

As can be seen from Table 8, the BW-K model has the optimal performance among best single model and simple average model in all households. On average, the BW-K model had a 0.97% performance improvement over the optimal single model and a 0.82% performance improvement over the simple average model. For household #1004, The performance of BW-K model is similar to SA-K model, but inferior to WA-K model. This is due to the good performance of the random forest model and the GBRT model in the test set, which perform poorly in the training set. Looking at other household results, it can be seen that not all the combined models are superior to the best single models, however, BW-K model performs well in our case study, which also explains the robustness of BW-K model in scenarios with high uncertainty and volatility.

7. Conclusions and future works

This paper proposes a method for transforming multiple quantile forecasts into a combined density forecast. Different statistic techniques are applied to generate high-quality quantile forecasting results. KDE is used to transform quantile forecasts into density forecasts. Perturbation search is designed to find optimal weights for model combination. CRPS and quantile loss function are introduced to evaluate forecasting performance. In the case studies, the BW-K model can achieve better performance than the best individual model and simple average model. At the same time, we also verify the influence of different kernel functions in KDE. The results indicate that there is almost no impact on the forecasting result among different kernel functions. Therefore, the proposed method owns robustness to the selection of the kernel function.

The research aims to propose a practical load probability density forecasting method by transforming and combining quantile forecasts. Future work can be extended to the following three areas:

1. **More forecasting scenarios:** Due to the different load characteristics in different regions, it is necessary to add more cases to further verify the practicability and effectiveness of the method;
2. **More elaborate segmentation methods:** In this paper, all combined works are done at the model level. However, the weights can act on quantile or each time point. The inertia and correction of the load in distribution may be better considered if we determined weights for each quantile. Similarly, we can set different weights for seasons or days and nights to take advantage of different models at different times.

Table 8
Average CRPS of residential smart meter data in Ireland for each individual model and combined model in the test set. The Gaussian kernel function is used for KDE.

Meter ID	Models										
	RF	RNN	GBRT	LR	LGBM	SA-E	WA-E	BW-E	SA-K	WA-K	BW-K
#1003	221.78	205.76	221.66	201.36	200.18	206.77	206.24	200.24	201.98	201.61	**197.34**
#1004	349.65	354.95	347.49	373.74	347.55	344.67	344.56	345.37	344.09	**343.99**	344.06
#1009	284.47	287.93	272.20	279.37	266.93	272.83	272.55	267.56	266.62	266.45	**264.41**
#1013	99.48	97.90	96.96	95.81	92.99	94.64	94.59	93.15	93.42	93.39	**92.62**
#1015	144.11	146.35	144.30	148.66	142.70	142.23	142.22	142.10	141.31	141.29	**141.16**

3. **Wider range of applications:** Since the methods used in this study are common mathematical methods, they are not limited to short-term load forecasting studies and can be applied in other fields. At present, the combining method faces problems such as long training and tuning time, hard to perform parameter self-updating and so on. In the future, further research is needed to cope with complex changes.

CRediT authorship contribution statement

Shu Zhang: Methodology, Software, Data curation, Visualization, Writing - original draft. **Yi Wang:** Conceptualization, Formal analysis, Visualization, Writing - review & editing. **Yutian Zhang:** Investigation, Validation. **Dan Wang:** Resources, Project administration. **Ning Zhang:** Writing - review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors 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 work was supported in part by International (Regional) Joint Research Project of National Natural Science Foundation of China (No. 71961137004) and Scientific & technical project of State Grid Research and application of operation platform for ubiquitous power Internet of things.

References

[1] 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.
[2] van der Meer D, Widén J, Munkhammar J. Review on probabilistic forecasting of photovoltaic power production and electricity consumption. Renew Sustain Energy Rev 2018;81:1484–512. http://dx.doi.org/10.1016/j.rser.2017.05.212, URL http://www.sciencedirect.com/science/article/pii/S1364032117308523.
[3] Irwin GW, Monteith W, Beattie WC. Statistical electricity demand modelling from consumer billing data. IEE Proc C - Gener Transm Distrib 1986;133(6):328–35. http://dx.doi.org/10.1049/ip-c.1986.0048.
[4] Herman R, Kritzinger J. The statistical description of grouped domestic electrical load currents. Electr Power Syst Res 1993;27(1):43–8. http://dx.doi.org/10.1016/0378-7796(93)90058-M, URL http://www.sciencedirect.com/science/article/pii/037877969390058M.
[5] Ghosh AK, Lubkeman DL, Downey MJ, Jones RH. Distribution circuit state estimation using a probabilistic approach. IEEE Trans Power Syst 1997;12(1):45–51. http://dx.doi.org/10.1109/59.574922.
[6] Heunis SW, Herman R. A probabilistic model for residential consumer loads. IEEE Trans Power Syst 2002;17(3):621–5. http://dx.doi.org/10.1109/TPWRS.2002.800901.
[7] Singh R, Pal BC, Jabr RA. Statistical representation of distribution system loads using Gaussian mixture model. IEEE Trans Power Syst 2010;25(1):29–37. http://dx.doi.org/10.1109/TPWRS.2009.2030271.
[8] Golestaneh F, Pinson P, Gooi HB. Very short-term nonparametric probabilistic forecasting of renewable energy generation— With application to solar energy. IEEE Trans Power Syst 2016;31(5):3850–63. http://dx.doi.org/10.1109/TPWRS.2015.2502423.
[9] Li T, Wang Y, Zhang N. Combining probability density forecasts for power electrical loads. IEEE Trans Smart Grid 2019. 1–1. http://dx.doi.org/10.1109/TSG.2019.2942024.
[10] 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.
[11] Yang Y, Li S, Li W, Qu M. Power load probability density forecasting using Gaussian process quantile regression. Appl Energy 2018;213:499–509. http://dx.doi.org/10.1016/j.apenergy.2017.11.035, URL http://www.sciencedirect.com/science/article/pii/S0306261917316100.
[12] 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, URL http://www.sciencedirect.com/science/article/pii/S0169207016000133.
[13] Gaillard P, Goude Y, Nedellec R. Additive models and robust aggregation for gefcom2014 probabilistic electric load and electricity price forecasting. Int J Forecast 2016;32(3):1038–50. http://dx.doi.org/10.1016/j.ijforecast.2015.12.001.
[14] Lebotsa ME, Sigauke C, Bere A, Fildes R, Boylan JE. Short term electricity demand forecasting using partially linear additive quantile regression with an application to the unit commitment problem. Appl Energy 2018;222:104–18. http://dx.doi.org/10.1016/j.apenergy.2018.03.155, URL http://www.sciencedirect.com/science/article/pii/S030626191830504X.
[15] Ben Taieb S, Huser R, Hyndman RJ, Genton MG. Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression. IEEE Trans Smart Grid 2016;7(5):2448–55. http://dx.doi.org/10.1109/TSG.2016.2527820.
[16] Meinshausen, Nicolai, Chridgeway SME. Quantile regression forests. J Mach Learn Res 2006;7(2):983–99.
[17] Wang Y, Gan D, Sun M, Zhang N, Lu Z, Kang C. Probabilistic individual load forecasting using pinball loss guided LSTM. Appl Energy 2019;235:10–20. http://dx.doi.org/10.1016/j.apenergy.2018.10.078, URL http://www.sciencedirect.com/science/article/pii/S0306261918316465.
[18] Gan D, Wang Y, Yang S, Kang C. Embedding based quantile regression neural network for probabilistic load forecasting. J. Modern Power Syst. Clean Energy 2018;6(2):244–54. http://dx.doi.org/10.1007/s40565-018-0380-x.
[19] Zhang W, Quan H, Srinivasan D. An improved quantile regression neural network for probabilistic load forecasting. IEEE Trans Smart Grid 2019;10(4):4425–34. http://dx.doi.org/10.1109/TSG.2018.2859749.
[20] He Y, Qin Y, Wang S, Wang X, Wang C. Electricity consumption probability density forecasting method based on LASSO-quantile regression neural network. Appl Energy 2019;233–234:565–75. http://dx.doi.org/10.1016/j.apenergy.2018.10.061, URL http://www.sciencedirect.com/science/article/pii/S0306261918316301.
[21] Rosenblatt M. Remarks on some nonparametric estimates of a density function. Ann Math Stat 1956;27(3):832–7.
[22] Parzen E. On estimation of a probability density function and mode. Ann Math Stat 1962;33(3):1065–76. http://dx.doi.org/10.1214/aoms/1177704472.
[23] Haben S, Giasemidis G. A hybrid model of kernel density estimation and quantile regression for gefcom2014 probabilistic load forecasting. Int J Forecast 2016;32(3):1017–22. http://dx.doi.org/10.1016/j.ijforecast.2015.11.004, URL http://www.sciencedirect.com/science/article/pii/S0169207015001399.
[24] He Y, Liu R, Li H, Wang S, Lu X. Short-term power load probability density forecasting method using kernel-based support vector quantile regression and copula theory. Appl Energy 2017;185:254–66. http://dx.doi.org/10.1016/j.apenergy.2016.10.079, URL http://www.sciencedirect.com/science/article/pii/S0306261916315239.
[25] Cao Z, Wan C, Zhang Z, Li F, Song Y. Hybrid ensemble deep learning for deterministic and probabilistic low-voltage load forecasting. IEEE Trans Power Syst 2019. 1–1. http://dx.doi.org/10.1109/TPWRS.2019.2946701.
[26] Pierro M, Bucci F, Felice MD, Maggioni E, Moser D, Perotto A, et al. Multi-model ensemble for day ahead prediction of photovoltaic power generation. Sol Energy 2016;134:132–46. http://dx.doi.org/10.1016/j.solener.2016.04.040, URL http://www.sciencedirect.com/science/article/pii/S0038092X16300731.
[27] 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.
[28] Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. J Amer Statist Assoc 2007;102(477):359–78.

[29] Jolliffe IT, Stephenson DB. Comments on "Discussion of Verification Concepts in Forecast Verification: A Practitioner's Guide in Atmospheric Science". Weather Forecast 2005;20(5):796–800. http://dx.doi.org/10.1175/WAF877.1.

[30] Thorey J, Chaussin C, Mallet V. Ensemble forecast of photovoltaic power with online CRPS learning. Int J Forecast 2018;34(4):762–73. http://dx.doi.org/10.1016/j.ijforecast.2018.05.007.

[31] Cho K, van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar: Association for Computational Linguistics; 2014, p. 1724–34. http://dx.doi.org/10.3115/v1/D14-1179, URL https://www.aclweb.org/anthology/D14-1179.

[32] Wand MP, Jones MC. Kernel smoothing. Chapman and Hall/CRC; 1994.

[33] He Y, Zheng Y. Short-term power load probability density forecasting based on yeo-johnson transformation quantile regression and Gaussian kernel function. Energy 2018;154:143–56. http://dx.doi.org/10.1016/j.energy.2018.04.072, URL http://www.sciencedirect.com/science/article/pii/S0360544218306790.

[34] He Y, Qin Y, Lei X, Feng N. A study on short-term power load probability density forecasting considering wind power effects. Int J Electr Power Energy Syst 2019;113:502–14. http://dx.doi.org/10.1016/j.ijepes.2019.05.063.

[35] He Y, Xu Q, Wan J, Yang S. Short-term power load probability density forecasting based on quantile regression neural network and triangle kernel function. Energy 2016;114:498–512. http://dx.doi.org/10.1016/j.energy.2016.08.023, URL http://www.sciencedirect.com/science/article/pii/S0360544216311264.

[36] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. Tensorflow: A system for large-scale machine learning. In: 12th USENIX symposium on operating systems design and implementation (OSDI 16). Savannah, GA: USENIX Association; 2016, p. 265–83, URL https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi.

[37] Crowther PS, Cox RJ. A method for optimal division of data sets for use in neural networks. In: International conference on knowledge-based and intelligent information and engineering systems. Springer; 2005, p. 1–7.

[38] Morrison R, Bryant C, Terejanu G, Miki K, Prudhomme S. Optimal data split methodology for model validation. 2011, arXiv preprint arXiv:1108.6043.

[39] Bergmeir C, Benítez JM. On the use of cross-validation for time series predictor evaluation. Inform Sci 2012;191:192–213.