Combining Probabilistic Load Forecasts

Yi Wang, *Student Member, IEEE*, Ning Zhang, *Senior Member, IEEE*, Yushi Tan,
Tao Hong, *Senior Member, IEEE*, Daniel S. Kirschen, *Fellow, IEEE*, and Chongqing Kang, *Fellow, IEEE*

Abstract—Probabilistic load forecasts provide comprehensive information about future load uncertainties. In recent years, many methodologies and techniques have been proposed for probabilistic load forecasting. Forecast combination, a widely recognized best practice in point forecasting literature, has never been formally adopted to combine probabilistic load forecasts. This paper proposes a constrained quantile regression averaging (CQRA) method to create an improved ensemble from several individual probabilistic forecasts. We formulate the CQRA parameter estimation problem as a linear program with the objective of minimizing the pinball loss and the constraints that the parameters are nonnegative and summing up to one. We demonstrate the effectiveness of the proposed method using two publicly available datasets, the ISO New England data and Irish smart meter data. Comparing with the best individual probabilistic forecast, the ensemble can reduce the pinball score by 4.39% on average. The proposed ensemble also demonstrates superior performance over nine other benchmark ensembles.

Index Terms—Probabilistic load forecasting, quantile regression, pinball loss function, ensemble method, linear programming, forecasts combination.

NOMENCLATURE

Indices

n	Index of probabilistic forecasting methods
q	Index of quantiles
t	Index of time periods.

Sets

N	Set of forecasting methods
Q	Set of quantiles
T	Set of time periods.

Functions

$f_{n,q}(\cdot)$	The n-th model for the q-th quantile
$f_{e,q}(\cdot)$	The ensemble model for the q-th quantile
$f_n(\cdot)$	The n-th point forecast model.

Variables

$\boldsymbol{\beta}_{Aq}$	Parameters of (C)QRA_A model for the q-th quantile
$\boldsymbol{\beta}_{Tq}$	Parameters of (C)QRA_T model for the q-th quantile
$\boldsymbol{\beta}_{Eq}$	Parameters of (C)QRA_E model for the q-th quantile
$\boldsymbol{\omega}$	Vector form of the weights of the base methods
$\boldsymbol{\omega}_q$	Vector form of the weights for the q-th quantile
$\hat{y}_{n,t,q}$	The forecasted q-th quantile of the n-th method at time t
$\hat{y}_{t,q}$	The forecasted q-th quantile of the ensemble method at time t
ω_n	Weight of the n-th method
$\omega_{n,q}$	Weight of the n-th method for the q-th quantile
$\mathbf{S}_{At}$	All quantiles of the N models
$\mathbf{S}_{Et}$	Average quantiles of the N models
$\mathbf{S}_{Tt,q}$	Targeted q-th quantiles of the N models
$\mathbf{S}_t$	Sorted quantiles at time t
$\mathbf{W}_{n,q}$	Parameters of the n-th model for the q-th quantile
$\mathbf{X}_{n,t}$	Inputs of the n-th model at time t
L	Total pinball loss of the ensemble method
$L_{n,t,q}$	Pinball loss of the n-th method at time t for the q-th quantile
L_n	Total pinball loss of the n-th method
$L_{t,q}$	Pinball loss of the ensemble method at time t for the q-th quantile
y_t	Actual load at time t.

Y. Wang, N. Zhang, and C. Kang are with the State Key Laboratory of Power Systems, Department of Electrical Engineering, Tsinghua University, Beijing 100084, China (e-mail: cqkang@tsinghua.edu.cn).

Y. Tan and D. S. Kirschen are with the Department of Electrical Engineering, University of Washington, Seattle, WA 98195-2500 USA.

T. Hong is with the Department of Systems Engineering and Engineering Management, University of North Carolina at Charlotte, Charlotte, NC 28223 USA.

I. INTRODUCTION

THE WAVE of embarking data analytics is changing the traditional practices of business operations and planning across many sectors including the power industry. The applications of smart meter data analytics are expanding to many corners of the power delivery systems, such as load analysis, load forecasting, and load management [1]. As an integral part of power system planing and operations, load forecasting has been extensively studied in [2]. The deployment of smart grid technologies, popularity of big data, and needs of describing future uncertainties are demanding highly accurate load forecasting models and even probabilistic load forecasts [3].

Majority of the load forecasting literature has been on point forecasting, which generates single-valued forecast or expected load for each timestamp in the future. The IEEE Working Group on Energy Forecasting organized the Global

Energy Forecasting Competition 2012 (GEFCom2012), which had a track devoted to point load forecasting. Many novel ideas and methodologies were highlighted by the winning teams [4]. The competition data was then published and used in many later case studies for benchmarking purposes [5], [6]. Some recent advancements in point load forecasting methodologies were driven by the usage of various data sources, such as smart meter data [7], [8], calendar data [9], [10], and non-temperature weather data, such as humidity [11] and wind speed [12]. The concerns about data quality also led to research in robustness benchmark, anomaly detection, and data cleansing [13], [14] for load forecasting.

Probabilistic forecasts can provide more comprehensive information about future uncertainties that what point forecasts can do [2], [15]. Many power companies rely on probabilistic load forecasts for systems planning [16]. Producing forecasts in quantiles was the theme of GEFCom2014 [17]. In the probabilistic load forecasting track, Gailand *et al.* [18] used a quantile generalized additive model (quantGAM), a hybrid of quantile regression and generalized additive models. Dordonnat *et al.* [19] first developed a point forecasting model based on semi-parametric regression, and then fed the model with different temperature scenarios to generate probabilistic forecasts. Another winning team, Jingrui Xie, took a similar strategy by feeding temperature scenarios to a multiple linear regression based point forecasting model [20]. Xie and Hong [21] further compared three temperature scenario generation methods for probabilistic load forecasting. Researchers have also developed other means to generate probabilistic forecasts, such as residual simulation [22], combining point forecasts [23], and using probabilistic forecasting techniques such as quantile regression [24], [25]. Model selection for probabilistic load forecasting was investigated in [26]. Some recent work also considered behind-the-meter solar generation in probabilistic net load forecasting [27].

Probabilistic load forecasting has also been applied to individual loads at household or building level to capture uncertainty. Arora and Taylor [28] captured the uncertainties on individual residential load profiles using both kernel density and conditional kernel density estimation methods. They also demonstrated that conditional kernel density can capture the seasonality and describe kernel bandwidth selection and boundary correction methods. Taieb *et al.* [24] combined gradient boosting and additive models in a single quantile regression method. Kou and Gao [29] proposed a sparse heteroscedastic forecasting method based on a Gaussian process for day-ahead forecasting in energy-intensive enterprises.

It is widely agreed that no individual forecasting method is the best for all datasets. Combining different forecasts usually reduces the overall risk of making a poor model selection. Forecast combination or ensemble methods can be classified into homogeneous ensemble methods and heterogeneous ensemble methods [30]. The former uses the same algorithm with diverse input data, input features, or output targets, while the latter combines several forecasts with the hope that diversity can help improve the results.

Both methods have been adopted in point load forecasting. Nowotarski *et al.* [31] took the homogeneous method, showing that ensembles of 8 sister load forecasts outperformed the best individual point forecasts. Li *et al.* [32] developed a homogeneous ensemble method by varying the mother wavelet and the number of decompositions of the wavelet transformation. They then combined the forecasts using partial least squares regression. At NPower Forecasting Challenge 2015, the BigDEAL team, who won a top 3 place, used a heterogeneous ensemble by combining independent forecasts from various techniques, such as multiple linear regression, autoregressive integrated moving average, artificial neural networks, and random forecasts [33]. Dudek combined 10 forecasts from different models, including *k*-means, exponential smoothing, and neural networks using simple averaging and weighted averaging, where the weights were determined by the performance of the individual models [34].

Some combination methods can also be applied to different individual probabilistic forecasts to produce another probabilistic forecast. Hall and Mitchell [35] chose to minimize the Kullback-Leibler distance between the forecast and true but unknown densities. Clements and Harvey [36] guided the search for optimal weights with the logarithmic probability score (LPS) and use the maximum likelihood (ML) method to solve the weights. Jore *et al.* [37] used recursive weights to combine the vector autoregressive (VAR)- and autoregressive (AR)-based density forecasts. Combining probabilistic forecasts is still challenging because the problem is usually formulated as a nonlinear and non-convex optimization problem, so that global optimality cannot be guaranteed and the combined results may be worse than individual forecasts. Bracale *et al.* [38] provided a probabilistic competitive ensemble method combining the Bayesian model, Markov chain model, and quantile regression method, where the weights assigned to each model are determined by minimizing the Continuous Ranked Probability Score (CRPS) of the overall ensemble.

The literature of combining probabilistic load forecasts is still quite limited. Mangalova and Shesterneva [39] used Nadaraya-Watson estimators to combine multiple density forecasts in a sequential fashion, which is similar to the procedure of model generation in the bagging-based ensemble method. Haben and Giasemidis [40] used different weights for the kernel density estimation method and the quantile regression methods to achieve minimum pinball scores. Nevertheless, none of them addressed how to determine the optimal weights for more than two probabilistic forecasts.

Since many probabilistic forecasts characterize future uncertainties in the form of quantiles, combining these quantile forecasts to improve the skill of the ensemble would be quite meaningful in practice. In this paper, we contribute to the probabilistic load forecasting literature by proposing a constrained quantile regression averaging (CQRA) method for quantile forecast combination. For parameter estimation, we formulate the proposed CQRA method as a linear programming (LP) problem that minimizes the pinball loss, of which the solution includes the optimal weights for the individual probabilistic forecasts. We construct case studies based on two

publicly available datasets: the zonal and system level loads from ISO New England (ISO-NE) and the household-level loads from the Commission for Energy Regulation (CER) in Ireland. When using the proposed method to combine forecasts at the target quantile, we can obtain an ensemble that outperforms all individual probabilistic forecasts as well as nine other benchmarks most of the time.

The rest of this paper is organized as follows. Section II introduces the probabilistic load forecast combination problem and identifies several key challenges. Section III introduces the techniques that we are going to use to generate individual probabilistic forecasts. Section IV derives an LP formulation of the probabilistic forecast combination problem. Section V introduces evaluation criteria to quantify the performance of probabilistic forecasts and lists several benchmarks for comparison. Section VI presents the case study settings, forecast results and comparison. Section VII conclusions the paper with an outlook of future research.

II. PROBLEM STATEMENT

Forecasting combination refers to the integration of a series of forecasting models to formulate the final forecasting model [30]. The combination process can be divided into two stages: 1) generating a set of forecasting models; 2) integrate these models. This paper endeavors to combine various probabilistic forecasts to provide a improved quantile forecast. Two issues corresponding to the two processes should be addressed:

1) Model generating: how to generate different quantile probabilistic forecasts with higher diversity about future uncertainties?
2) Model ensemble: how to determine the weights of the individual models to achieve the optimal combination?

The first issue will be addressed in Section III by using three typical quantile regression models, and the second issue will be addressed in Section IV by formulating an optimization problem.

III. PROBABILISTIC LOAD FORECASTING MODEL GENERATION

In this section, both homogeneous and heterogeneous methods are used to obtain various basic quantile probabilistic load forecasting models. Density or distribution forecast results can also be converted to the form of quantiles. Since the application of basic forecasting models is not the main consideration of this paper, we only introduce three typical regression models for quantile probabilistic forecasts. We obtain different forecasts by varying the hyperparameters and input data.

Quantile regression itself can be formulated as an optimization problem to minimize the pinball loss. The pinball loss is a comprehensive index to evaluate the reliability, sharpness, and calibration of the forecasts. It was used as the error measure in GEFCom2014 and GEFCom2017. It is defined for any quantile $q \in (0, 1)$ through a weighted absolute error:

$$L_{n,t,q}(\hat{y}_{n,t,q}, y_t) = \begin{cases} (1-q)(\hat{y}_{n,t,q} - y_t) & \hat{y}_{n,t,q} \geq y_t \\ q(y_t - \hat{y}_{n,t,q}) & \hat{y}_{n,t,q} < y_t. \end{cases} \quad (1)$$

For a certain regression model, its parameters can be optimized by the following optimization problem:

$$\mathbf{W}_{n,q} = \arg \min_{\mathbf{W}_{n,q}} \sum_{t \in T} \sum_{q \in Q} L_{n,t,q}(f_{n,q}(\mathbf{X}_{n,t}, \mathbf{W}_{n,q}), y_t). \quad (2)$$

The regression functions $f_{n,q}(\cdot)$ are distinct for different quantile regression models, such as artificial neural network (ANN), gradient boosting regression tree (GBRT), and random forests (RF). The integration of these regression methods and the pinball loss function in load forecasting is briefly introduced in the following paragraphs.

A. Quantile Regression Neural Network (QRNN)

For linear regression, the parameters $\mathbf{W}_{n,q}$ include the slopes $\boldsymbol{\beta}_{n,q}$ and intercept $b_{n,q}$:

$$\hat{y}_t = f_n(\mathbf{X}_{n,t}, \mathbf{W}_{n,q}) = \boldsymbol{\beta}_{n,q}\mathbf{X}_{n,t} + b_{n,q}. \quad (3)$$

Similarly, for ANN with a hidden layer, the parameters $\mathbf{W}_{n,q}$ include the hidden layer weights $\mathbf{W}_{n,q}^{(h)}$, the hidden layer biases $\mathbf{b}_{n,q}^{(h)}$, the output weights $\mathbf{W}_{n,q}^{(o)}$, and the output biases $b_{n,q}^{(o)}$. Thus, the output of the hidden layer $\mathbf{y}_t^{(h)}$ can be calculated as follows:

$$\mathbf{y}_t^{(h)} = g_n^{(h)}\left(\mathbf{W}_{n,q}^{(h)}\mathbf{X}_{n,t} + \mathbf{b}_{n,q}^{(h)}\right), \quad (4)$$

where $g_n^{(h)}(\cdot)$ denotes the activation function of the hidden layer. Choices of activation function include sigmoid function, tanh function, and rectified linear unit (ReLU). We choose the sigmoid function in this paper.

The final output, i.e., the forecast value, is:

$$\hat{y}_t = g_n^{(o)}\left(\mathbf{W}_{n,q}^{(o)}\mathbf{y}_t^{(h)} + b_{n,q}^{(o)}\right), \quad (5)$$

where $g_n^{(o)}(\cdot)$ denotes the activation function of the output layer.

To reduce the risk of overfitting the ANN, a weight decay regularization term that penalizes large weights is added to the total pinball loss in Eq. (2).

$$\mathbf{W}_{n,q} = \arg \min_{\mathbf{W}_{n,q}} \sum_{t \in T} E\{L_{n,t,q}(\hat{y}_t, y_t)\} + \lambda \left\|\mathbf{W}_{n,q}^{(h)}\right\|, \quad (6)$$

where λ is a positive constant and represents the proportion of the weight decay term.

For traditional ANN with a mean square error (MSE) loss function, which is differentiable everywhere, the back-propagation algorithm can be used to search the weights and biases. For a QRNN with a pinball loss function, which is only partially differentiable, an approximation function that is differentiable everywhere can be used to replace the pinball loss function [41]. The number of neurons in the input layer is determined by the dimensions of the input data, and a different number of neurons in the hidden layer results in a different QRNN model. This number is set to 4, 5, 6, 7, and 8 to train five different QRNN models for the ensemble process.

B. Quantile Regression Random Forests (QRRF)

Random Forests (RF) is a decision tree (classification and regression tree, CART)-based learning method. Each decision tree is trained using a bootstrap sampled subset of the original training dataset. N_{RF} denotes the number of trees to grow and M_{RF} the number of variables randomly sampled as candidates at each split. Thus, the parameters $\mathbf{W}_{n,q}$ include random parameter $\mathbf{W}_{n,q,n_{RF}}$ for each individual decision tree $T(\mathbf{W}_{n,q,n_{RF}})$. The parameters for each tree can be obtained through the M_{RF} training data, where the training method is the same as the CART training method. Each individual decision tree provides a forecast result:

$$\hat{y}_{t,n_{RF}} = T\big(\mathbf{W}_{n,q,n_{RF}}, \mathbf{X}_{n,t}\big). \tag{7}$$

Then, the forecast of RF is:

$$\hat{y}_t = \frac{1}{N_{RF}} \sum_{n_{RF}=1}^{N_{RF}} \hat{y}_{t,n_{RF}}. \tag{8}$$

A traditional RF attempts to reduce the absolute error by approximating the conditional mean. Because a RF can provide multiple estimates of y_t, a QRRF can be easily implemented by computing the distribution function and quantiles based on these estimates of $\hat{y}_{t,n_{RF}}$. While RF is essentially an ensemble regression method, it is used as a base probabilistic load forecasting model in this paper.

To guarantee that every input row is predicted at least a few times, the number of trees to grow N_{RF} should not be too small. $N_{RF} = 500$ is used in this paper. However, the number of variables randomly sampled from dataset M_{RF} is set to 1/2, 1/3, 1/4, and 1/5 of the size of the whole training dataset to train four different QRRF models for the ensemble process.

C. Quantile Regression Gradient Boosting (QRGB)

The gradient boosting decision tree (GBDT) method is also a decision-tree-based ensemble regression method. In contrast to the RF, GBDT adds a new decision tree to fit the residuals from previously generated trees. The final regression model is trained in an iterative manner. We denote M_{GB} as the total number of terminal nodes of the learned tree and N_{GB} as the number of iterations. Thus, the parameters $\mathbf{W}_{n,q}$ include a random parameter $\mathbf{W}_{n,q,n_{GB}}$ for each individual decision tree $f_{n,n_{GB}}(\mathbf{W}_{n,q,n_{GB}})$. For the n_{GB}-th iteration,

$$f_n^{n_{GB}}(\mathbf{X}_t) = f_n^{n_{GB}-1}\big(\mathbf{X}_{n,t}\big)$$
$$+ \lambda \sum_{m_{GB}=1}^{M_{GB}} \gamma_{m_{GB},n_{GB}} 1\big(\mathbf{X}_{n,t} \in R_{m_{GB},n_{GB}}\big), \tag{9}$$

where λ denotes the learning rate of the gradient boosting tree; $\{R_{m_{GB},n_{GB}}\}_1^{M_{GB}}$ denotes the M_{GB} disjoint regions at the n_{GB}-th iteration; $\gamma_{m_{GB},n_{GB}}$ denotes the optimal terminal node estimates and can be calculated as follows:

$$\gamma_{m_{GB},n_{GB}} = \arg\min_{\gamma} \sum_{t \in T}$$
$$L_{n,t,q}\big(f_n^{n_{GB}-1}\big(\mathbf{X}_{n,t}, \mathbf{W}_{n,q,n_{GB}-1}\big) + \gamma, y_t\big). \tag{10}$$

In this way, each iteration attempts to minimize the previous residuals. More details about the derivation of GBDT can be found in [42]. One of the main hyperparameters of the QRGB is the number of iterations, i.e., the total number of trees to fit, N_{GB}. We vary N_{GB} from 70 to 100 with an interval of 10 to obtain four QRGB models for the ensemble process.

D. Summary

A total of thirteen quantile regression models are generated:
1) Five QRNN models with 4, 5, 6, 7, or 8 neurons in the hidden layer;
2) Four QRRF models where the number of variables randomly sampled from M_{RF} is set to 1/2, 1/3, 1/4, or 1/5 of the whole training dataset;
3) Four QRGB models where the number of trees to fit, N_{GB}, is set to 70, 80, 90, or 100.

These thirteen quantile regression models, both homogeneous models with different hyperparameters and heterogeneous models, including ANN, RF, and GBDT, constitute a comprehensive model set.

IV. CQRA BASED MODEL COMBINATION

Different indices can be used to quantify the performance of probabilistic load forecasting. For example, reliability and sharpness describe how close and tightly the predicted distribution is to the actual one, respectively; resolution describes how much the predicted interval varies over time [2]. Instead of focusing on a single aspect, pinball loss provides a comprehensive score for the performance of probabilistic forecasts and is consistent with the objective function of quantile regression. Thus, the pinball loss is used as a guide for the integration of multiple models.

A. Problem Formulation

If the weights of each individual model for all the quantiles are identical, the forecast combination model for quantile q can be formulated as follows:

$$f_{e,q}(\mathbf{X}_{n,t}, \boldsymbol{\omega}) = \sum_{n=1}^{N} \omega_n f_{n,q}\big(\mathbf{X}_{n,t}, \mathbf{W}_{n,q}\big). \tag{11}$$

The weights $\boldsymbol{\omega}$ are estimated by solving the following optimization problem:

$$\hat{\boldsymbol{\omega}} = \arg\min_{\boldsymbol{\omega}} \quad \sum_{q \in Q} \sum_{t \in T} L_{n,t,q}\left(\sum_{n=1}^{N} \omega_n f_{n,q}(\mathbf{X}_{n,t}, \mathbf{W}_{n,q}), y_t\right)$$

$$\text{s.t.} \quad \sum_{n=1}^{N} \omega_n = 1$$
$$\omega_n \geq 0, \quad \forall n \in \{1, \dots, N\}. \tag{12}$$

By contrast, if the weights of each individual model for the different quantiles are different, the forecast combination method is formulated as:

$$f_{e,q}(\mathbf{X}_{n,t}, \boldsymbol{\omega}_q) = \sum_{n=1}^{N} \omega_{n,q} f_{n,q}\big(\mathbf{X}_{n,t}, \mathbf{W}_{n,q}\big). \tag{13}$$
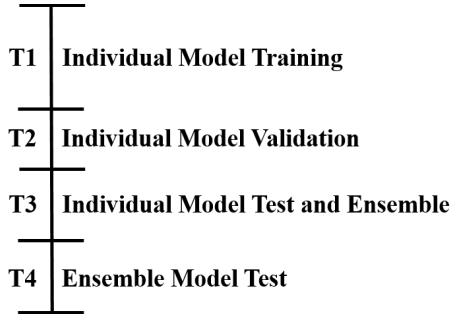
Fig. 1. Time series splitting for model training, validation, integration, and testing.

In the following, we focus on the forecast combination method for each quantile. For each quantile, the determination of weights $\boldsymbol{\omega}$ is transformed into solving Q individual optimization problems. The q-th problem is:

$$\hat{\boldsymbol{\omega}}_q = \arg\min_{\boldsymbol{\omega}_q} \quad \sum_{t\in T} L_{n,t,q}\left(\sum_{n=1}^{N} \omega_{n,q}\, f_{n,q}(\mathbf{X}_{n,t}, \mathbf{W}_{n,q}), y_t\right)$$
$$\text{s.t.} \quad \sum_{n=1}^{N} \omega_{n,q} = 1,$$
$$\omega_{n,q} \geq 0, \quad \forall n \in \{1, \ldots, N\}. \tag{14}$$

B. Data Splitting

The naïve models shown in (12) and (14) have a high risk of overfitting. We split a validation dataset from the original training dataset to reduce the overfitting risk. Thus, the whole dataset is divided into four parts, as shown in Fig. 1. The first part T1 is used to train each of the models. The second part T2 is used to validate each model for hyperparameter tuning. The third part T3 is used to test each individual model, and the forecast results are used for the model combination in (12) or (14) such that the overfitting risk is reduced. The last part T4 is used to test the final combined forecasts. Since the solution methods for (12) and (14) would be the same, in the following, the algorithms for model integration are described by taking (12) as an example.

How to optimally split the time series for model training, validation, integration, and testing is beyond the scope of this paper, and therefore the time series splitting in this paper is roughly conducted based on the number of years. The optimal data split issue has been studied in [43], where the split ratio for model training and validation was investigated. For our problem, the data split is to determine three time breakpoints for T1/T2, T2/T3, and T3/T4. Grid search can be probably applied to search the optimal data split. We would like to leave it for our future work.

C. Model Combination

In this subsection, an algorithm to solve the optimization problem is proposed for model combination.

Since with a limited number of quantiles we cannot obtain a complete distribution, a certain quantile of the weighted sum of several quantile-based distributions cannot be calculated either.

In this paper, as proposed in [44], we simply estimate the q-th quantile of the weighted sum of several distributions by the weighted sum of the q-th quantiles of all the distributions:

$$\hat{y}_{t,q} \approx \sum_{n\in N} \omega_{n,q}\hat{y}_{n,t,q}. \tag{15}$$

Thus, the loss function in (12) can be rewritten as follows:

$$\hat{\boldsymbol{\omega}}_q = \arg\min_{\boldsymbol{\omega}_q} \sum_{t\in T} L_{t,q}(\hat{y}_{t,q}, y_t)$$
$$= \arg\min_{\boldsymbol{\omega}_q} \sum_{t\in T} \sum_{q\in Q} \max\{q(y_t - \hat{y}_{t,q}), (1-q)(\hat{y}_{t,q} - y_t)\}$$
$$\text{s.t.} \quad \hat{y}_{t,q} = \sum_{n\in N} \omega_{n,q}\hat{y}_{n,t,q}, \quad \sum_{n\in N} \omega_{n,q} = 1, \quad \omega_n \geq 0 \quad \forall n. \tag{16}$$

For the q-th quantile at time t, we introduce auxiliary decision variables $v_{t,q} = \max\{q(y_t - \hat{y}_{t,q}), (1 - q)(\hat{y}_{t,q} - y_t)\}$, so the problem in (16) can be transformed into:

$$\hat{\boldsymbol{\omega}}_q = \arg\min_{\boldsymbol{\omega}_q} \sum_{t\in T} v_{t,q}$$
$$\text{s.t.} \quad \hat{y}_{t,q} = \sum_{n\in N} \omega_{n,q}\hat{y}_{n,t,q}, \quad \sum_{n\in N} \omega_{n,q} = 1, \quad \omega_{n,q} \geq 0 \quad \forall n.$$
$$v_{t,q} \geq q(y_t - \hat{y}_{t,q}), \quad v_{t,q} \geq (1 - q)(\hat{y}_{t,q} - y_t)$$
$$\{v_{t,q} - q(y_t - \hat{y}_{t,q})\}\{v_{t,q} - (1 - q)(\hat{y}_{t,q} - y_t)\} = 0. \tag{17}$$

Without the last constraint in (17), the problem is an LP problem. We denote the model without the last constraint in (17) as the RLP model. Reduction to absurdity could be used to prove that the optimal solution of RLP is also the optimal solution of model (17).

Proof: If the optimal solution of RLP $[\boldsymbol{\omega}, \mathbf{v}]$ does not satisfy the last constraint, then there exists at least one quantile and time period that satisfies $v_{t,q} > q(y_t - \hat{y}_{t,q})$ and $v_{t,q} > (1 - q)(\hat{y}_{t,q} - y_t)$. Then, we can find another value $v'_{t,q} = v_{t,q} - \varepsilon$ that satisfies $v'_{t,q} = q(y_t - \hat{y}_{t,q}) > (1 - q)(\hat{y}_{t,q} - y_t)$ or $v'_{t,q} = (1 - q)(\hat{y}_{t,q} - y_t) > q(y_t - \hat{y}_{t,q})$, where ε is a positive value. Then, $[\boldsymbol{\omega}, \mathbf{v}']$ instead of $[\boldsymbol{\omega}, \mathbf{v}]$ is the optimal solution of RLP. ∎

Thus, model (14) can be transformed into an LP problem with approximation on the combined quantiles with T more variables and $2 \times T$ more constraints.

Remark 1: Quantile crossing is a well-known but still not well-addressed issue when quantile regression models fitted separately for different quantiles [45]. There are two possible ways to avoid quantile crossing in this context. The first one is to integrate the combination models for individual quantiles into one model. The objective function is to minimize the total pinball losses of all quantiles, with more constraints to avoid quantile crossing. The second one is to conduct naive rearrangement after all quantiles are obtained [40], [46]. In this paper, we choose naive rearrangement for three reasons: 1) The individually trained combination models work well in our case studies for both system-level load datasets and residential load datasets with very few time periods existing quantile crossing. The first integration strategy involves large amounts of variables and additional constraints and makes the optimization model much more complex and time-consuming.

It might not worth the effort of solving such a complicated model just to hedge against this few quantile crossings; 2) The integration strategy can only guarantee that there is no quantile crossing the ensemble stage. Quantile crossing could still happen in the test and prediction stage. 3) Naive rearrangement has very low computational burden as well as nice asymptotic properties [46].

Remark 2: Bayesian model averaging (BMA) is another effective way to combine different models to provide probabilistic forecast [47]. It uses posterior probability as the weight to conduct weighted average to individual models. It considers the model uncertainty and combines subjective information, model and data information by setting different prior distributions. The main challenges of the implementation of BMA include model selection, integral computation, individual model posterior probability calculation [48]. Our proposed combining method differs from BMA in different ways: 1) the proposed method determines the weights of individual methods under the guidance of pinball loss instead of the posterior probabilities; 2) some of the models that we are trying to combine do not have explicit probability structures for calculating likelihood functions. Therefore, the proposed method should be more universal; 3) the proposed method can be formulated as a linear programing problem and is of computational efficiency and accuracy; whereas posteriors sometimes present multi-modality which may be undesirable in decision making [49]; 4) the proposed formulation can be viewed as a special case of lasso regression [50] and therefore select predictors by enforcing sparsity. However, the candidate models for BMA need to be either carefully pre-screened or adaptively dropped in a heuristic way.

V. EVALUATION INDEXES AND COMPETING METHODS

In this section, the evaluation index pinball loss is used to quantify the performance of probabilistic load forecasts. In addition, nine different weight determination methods for model ensemble are introduced for comparison in the case studies.

A. Evaluation Index

As stated before, the probabilistic forecasts can be evaluated from different aspects including reliability, sharpness, and resolution [2]. The CRPS is a comprehensive index for the three aspects which is mainly used for density forecasts instead of quantiles. The discrete form of CRPS was proposed in [24]. Another index, pinball loss, is another comprehensive index that has been used in many previous studies [2] and global energy forecasting competitions. Thus, we use pinball to evaluate the final forecasts. It is calculated using (1). The average of all the quantiles' pinball loss gives the overall performance:

$$L = \frac{1}{|T_4| \times Q} \sum_{t \in T_4} \sum_{q \in Q} L_{t,q}(\hat{y}_{t,q}, y_t). \tag{18}$$

where $|T_4|$ denotes the length of the time period T_4.

B. Competing Methods

The nine competing methods consist of naive sorting method, median value based method, simple averaging, weighted averaging method, three quantile regression averaging (QRA) methods, and two CQRA methods.

1) Naïve Sorting (NS): With each forecasting model producing Q quantiles, a total of $N \times Q$ quantiles can be observed (in some sense) by N forecasting models. By sorting these observations by descending order, a new sequence $\mathbf{S}_t = \{S_{t,j}, j = [1, Q \times N]\}$ can be obtained. And therefore the q-th quantile is estimated as follows:

$$\hat{y}_{t,q} = S_{t,1+(q-1)N}. \tag{19}$$

2) Median Value (MED): The median value of the N q-th quantiles is selected as the final quantile:

$$\hat{y}_{t,q} = S_{t,1+(q-1)N+[N/2]}. \tag{20}$$

3) Simple Averaging (SA): The simple averaging strategy applies equal weights to different methods:

$$w_{n,q} = 1/N. \tag{21}$$

Then, the final combined forecasts are calculated according to Eq. (15).

4) Weighted Averaging (WA): The basic idea of the weighted averaging method is that methods with higher accuracy should be given higher weights:

$$w_{n,q} = \frac{\frac{1}{L_{n,q}}}{\sum_{n \in N} \frac{1}{L_{n,q}}}. \tag{22}$$

Similarly, the final combined forecasts are also calculated according to Eq. (15).

5) QRA-E: As stated above, N forecasting models produce $N \times Q$ quantiles. These quantiles can also be viewed as $N \times Q$ point forecasts which are denoted as $\mathbf{S}_{At}|_{1 \times (Q \times N)}$. Then, we can calculate the average quantiles $\mathbf{S}_{Et}|_{1 \times N}$, where the element of the average quantiles $\mathbf{S}_{Et,n}$ is calculated as follows:

$$S_{Et,n} = \frac{1}{Q} \sum_{i=Q \times (n-1)+1}^{Q \times n} S_{At,i}. \tag{23}$$

QRA-E yields new quantiles by applying linear quantile regression to the average of the Q quantiles $\mathbf{S}_{Et}$:

$$\hat{y}_{t,q} = \mathbf{S}_{Et} \boldsymbol{\beta}_{Eq}. \tag{24}$$

The parameters are estimated by minimizing the Pinball loss function:

$$\boldsymbol{\beta}_{Eq} = \arg\min_{\boldsymbol{\beta}_{Eq}} \sum_{t \in T} \sum_{q \in Q} L_{n,t}(\mathbf{S}_{At} \boldsymbol{\beta}_{Eq}, y_t). \tag{25}$$

6) QRA-A: Compared with QRA-E, instead of conducting QRA on the averaged quantiles, QRA-A yields new quantiles by applying linear quantile regression to all quantiles $\mathbf{S}_{At}$:

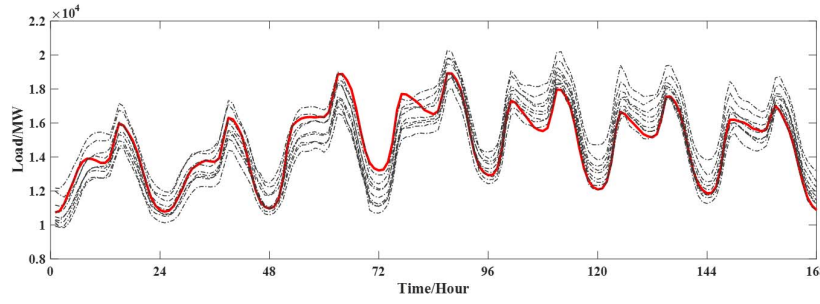$$\hat{y}_{t,q} = \mathbf{S}_{At} \boldsymbol{\beta}_{Aq}. \tag{26}$$

Fig. 2. Probabilistic load forecasts of the total load (SYS) of one week from Jan 2, 2016 to Jan 8, 2016, where the red line is the real values and the dotted line are forecasted quantiles.

TABLE I
COMPETING QRA METHODS

Constraints / Quantiles	With Constraints	Without Constraints
Averaged Quantiles	5) QRA-E	8) CQRA-E
All Quantiles	6) QRA-A	9) CQRA-A
Targeted Quantiles	7) QRA-T	CQRA-T (Proposed)

7) QRA-T: We can also select the targeted q-quantiles from $\mathbf{S}_{At,q}$ for QRA. The targeted quantiles $\mathbf{S}_{Tt,q}|_{1\times N}$ are selected as follows:

$$S_{Tt,q,n} = S_{At,Q\times(n-1)+q}. \quad (27)$$

Compared with QRA-A, a slight change of QRA-E is the regressors are the targeted q-th quantiles $\mathbf{S}_{Tt,q}$:

$$\hat{y}_{t,q} = \mathbf{S}_{Tt,q}\boldsymbol{\beta}_{Tq}. \quad (28)$$

It should be noted that the quantiles used for QRA are different and of different dimensions. Thus, the dimensions of the parameters (weights) of the regression model $\boldsymbol{\beta}_{Eq}$, $\boldsymbol{\beta}_{Aq}$, $\boldsymbol{\beta}_{Tq}$ are also different.

8) CQRA-E: Compared with QRA-E, the slight changes of CQRA-E are the added constraints:

$$\boldsymbol{\beta}_{Eq} = \arg\min_{\boldsymbol{\beta}_{Eq}} \sum_{t\in T}\sum_{q\in Q} L_{t,q}\big(\mathbf{S}_{Et}\boldsymbol{\beta}_{Eq}, y_t\big).$$

$$\text{s.t.} \quad \sum_{n\in N}\beta_{Eq,n} = 1, \quad \beta_{Eq,n} \geq 0 \quad \forall n. \quad (29)$$

9) CQRA-A: Compared with CQRA-E, CQRA-A conducts constrained regression on all the quantiles $\mathbf{S}_{At}$:

$$\boldsymbol{\beta}_{Aq} = \arg\min_{\boldsymbol{\beta}_{Aq}} \sum_{t\in T}\sum_{q\in Q} L_{t,q}\big(\mathbf{S}_t\boldsymbol{\beta}_{Aq}, y_t\big).$$

$$\text{s.t.} \quad \sum_{n\in N}\beta_{Aq,n} = 1, \quad \beta_{Aq,n} \geq 0 \quad \forall n. \quad (30)$$

Note that the proposed method in this paper can be called CQRA-T, because the targeted quantiles are used for regressors. The estimated weights for each quantile $\boldsymbol{\omega}_q$ is equal to the parameters $\boldsymbol{\beta}_{Tq}$ in CQRA-T model.

Table I summarized the competing QRA methods from two aspects: 1) whether there are constraints for the weights of the models; 2) which quantiles have been considered in the QRA method.

VI. CASE STUDIES

We conduct the case studies on both a zonal/system-level load dataset and an individual consumer load dataset on a standard PC with an Intel Core i7-4770MQ CPU running at 2.40 GHz and with 8.0 GB of RAM. The thirteen quantile regression models are implemented using the qrnn package [51], quantregforest package [52], and gbm package [53] in R version 3.3.2. The optimization model for the ensemble is implemented using YALMIP [54] in MATLAB R2016a.

A. Day-Ahead Zonal/System-Level Load Forecasting

1) Data Description: The case studies on the zonal/system-level load forecasting are conducted on an ISO-NE load dataset. ISO-NE, an independent regional transmission organization (RTO), serves six New England states: Connecticut (CT), Maine (ME), Massachusetts, New Hampshire (NH), Rhode Island (RI) and Vermont (VT), where Massachusetts consists of three zones: NEMASS, SEMASS, and WCMASS [55]. In this way, ISO-NE is divided into 8 zones. We validate the proposed method on these 8 zones and their total load, which is denoted as SYS. The hourly load data from 2013-1-1 to 2016-12-31, a total of four year data, are used. Specifically, the first two year of the data (T1 & T2) is used to train and validate the individual models; the third and fourth year of the data are used for the individual model combination (T3), and ensemble model test (T4), respectively.

2) Results: Fig. 2 provides the day-ahead probabilistic total (NC) load forecasting results of one week. The red line denotes the actual loads; the gray dotted lines denote the quantiles from 10% to 90% with intervals of 10%.

Table II provides the pinball losses of the different methods for the eight zones and the total load, where the best individual (BI) is the individual model with the best performance. Correspondingly, Fig. 3 shows the relative improvements compared with the BI model.

It can be seen that most of the combining models (simple averaging, weighted averaging, QRA, and the proposed optimal averaging) outperform the BI model. There is a special case where the two QRA models for VT has very slightly weak performance compared with BI model. It indicates that the combination method does not always provides better result than the individual methods but would provide improvements with higher credit. In addition, naive sorting has the worst performance because the there might be bias when determining
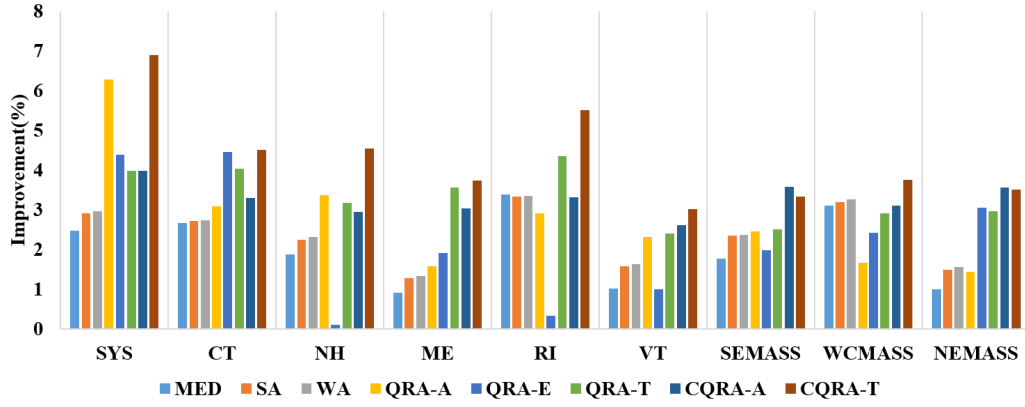
Fig. 3. Relative improvements of the different combination methods in different zones compared with the best individual (BI) model.

Quantiles \ Zones	CT	NH	ME	RI	VT	SEWASS	WCMASS	NEWASS	SYS
10-th	0.75	0.58	3.58	1.36	1.19	8.75	-1.49	4.56	2.82
20-th	0.41	1.92	2.20	2.25	1.02	7.78	3.15	4.05	-1.10
30-th	0.09	1.31	3.32	3.67	0.03	3.31	1.44	5.33	5.97
40-th	2.54	0.29	4.13	4.29	1.13	3.65	2.05	3.82	3.15
50-th	1.52	1.56	3.68	3.47	1.06	5.12	5.78	2.98	4.54
60-th	1.76	0.92	3.03	5.78	3.35	4.97	3.08	3.16	4.48
70-th	3.83	3.05	2.71	6.76	2.23	4.62	3.74	3.65	4.49
80-th	4.88	3.86	3.37	6.55	2.10	1.32	4.81	5.64	6.02
90-th	6.88	4.63	2.51	7.71	4.24	3.10	8.63	8.02	10.98

Fig. 4. Relative improvements (%) of the different combination methods for different quantiles in different zones.

Models \ Quantiles	10-th	20-th	30-th	40-th	50-th	60-th	70-th	80-th	90-th
#1	0	0	0	0.128	0.123	0	0.015	0	0.102
#2	0	0	0	0.177	0.022	0.236	0.154	0.004	0
#3	0.036	0	0	0.041	0.255	0	0.123	0.302	0
#4	0.385	0.444	0.281	0	0	0.030	0	0	0.068
#5	0.165	0	0	0.200	0.298	0.339	0.092	0	0.134
#6	0.037	0.093	0.537	0.264	0	0	0.000	0.251	0
#7	0	0.131	0	0.071	0	0	0.265	0.051	0.218
#8	0	0.207	0.152	0	0.158	0.003	0.350	0.133	0
#9	0.377	0.047	0.030	0.117	0.143	0.392	0	0.206	0.333
#10	0	0.078	0	0	0	0	0	0	0
#11	0	0	0	0	0	0	0	0.052	0.145
#12	0	0	0	0	0	0	0	0	0
#13	0	0	0	0	0	0	0	0	0

Fig. 5. Models that are selected for different quantiles for total load (SYS).

each quantiles according to the sorted sequence. The proposed combining method has the lowest pinball loss for all nine load profiles. The average improvement compared with the best individual is 4.39% which obviously verify the superiority of the proposed combining forecasts.

We also notice that the addition of the constraints to the method QRA-E (i.e., CQRA-E) so strongly worsen the performance of the combined forecasts. This is because the constraints $\sum_{n \in N} \beta_{Eq,n} = 1, \quad \beta_{Eq,n} \geq 0 \quad \forall n$ limit the combination into the a very small interval, the lower and upper bounds are the minimum and maximum excepted quantiles of the N models, respectively. QRA-E is quite similar to traditional QRA which applies quantile regression on the point forecasts. This is also the reason why traditional QRA has not constraints on the weights.

Fig. 4 depicts the relative improvements in our proposed method in different zones for different quantiles. The two points marked by the red circle show that not all the quantiles improve. There is no clear relationship between the improvement and the quantile.

In the process of model integration, some models are not shown be used to from the final forecasts. Fig. 5 provides the weights for the different models for the total load (SYS), respectively. Method #12 and #13 have been pruned for all quantiles owing to its poor performance. The number of models retained for different quantile ranges from 6 to 9. It should be noted that the pruned model does not necessarily have the worst performance. In addition, the weights for different models does not change smoothly with the change of quantiles. There may exist two reasons: 1) the proposed combining

Models \ Zones	SYS	CT	NH	ME	RI	VT	SEMASS	WCMASS	NEMASS
#1	0.102	0.144	0.231	0.015	0.001	0.355	0	0	0.196
#2	0	0	0	0.082	0.074	0.146	0.071	0	0
#3	0	0	0.031	0	0	0.079	0	0.196	0
#4	0.068	0	0.089	0.349	0	0	0.038	0	0
#5	0.134	0	0	0	0.272	0	0.199	0.318	0.199
#6	0	0	0.283	0.231	0.226	0.096	0	0	0.136
#7	0.218	0	0.058	0.058	0	0.082	0.166	0.218	0.049
#8	0	0.129	0.308	0.079	0.197	0	0.173	0.076	0.087
#9	0.333	0.341	0	0.185	0.021	0.243	0.290	0.192	0.333
#10	0	0	0	0	0	0	0	0	0
#11	0.145	0.267	0	0	0	0	0	0	0
#12	0	0	0	0	0.210	0	0	0	0
#13	0	0.119	0	0	0	0	0.062	0	0

Fig. 6. Models that are selected for the 90-th quantile for different zones.

method has constraints on the weights and results in sparsity of the weights; 2) the quantile regression models are trained individually for different individual quantiles and thus the performances of adjunct quantiles do not change smoothly. Fig. 6 gives the weights for the 90-th quantile for different zones in ISO-NE. The results shows that different models have been pruned for the 90-th quantile forecasting for different zones.

B. Hour-Ahead Residential Load Forecasting

1) Data Description: The case study on the individual consumer load forecasting is conducted on a CER Irish dataset, where ten residents are selected. The 30-min data are aggregated to one-hour load profiles from 2009-7-15 to 2010-12-31. The dataset division strategy for the training and testing of individual methods, and the ensemble is the same as that for the ISO-NE dataset.

TABLE II
PINBALL LOSSES OF THE INDIVIDUAL AND COMBINATION METHODS FOR DIFFERENT ZONES

| Zones Methods	SYS	CT	NH	ME	RI	VT	SEMASS	WCMASS	NEMASS
BI	288.563	81.478	27.216	18.146	21.756	12.426	42.307	41.939	63.685
NS	327.569	95.058	31.586	19.003	25.738	13.247	48.817	47.041	71.873
MED	281.607	79.359	26.713	17.981	21.044	12.300	41.570	40.676	63.048
SA	280.375	79.322	26.618	17.916	21.053	12.233	41.336	40.638	62.752
WA	280.266	79.306	26.600	17.908	21.049	12.227	41.329	40.616	62.706
QRA-E	276.417	77.995	27.184	17.806	21.683	12.303	41.484	40.949	61.793
QRA-A	271.519	79.037	26.330	17.864	21.140	12.145	41.295	41.252	62.783
QRA-T	277.487	78.313	26.380	17.523	20.847	12.135	41.271	40.752	61.849
CQRA-E	356.527	100.925	33.829	22.767	26.540	15.616	51.765	51.544	79.131
CQRA-A	277.510	78.870	26.437	17.610	21.059	12.109	**40.847**	40.672	**61.491**
CQRA-T	**269.953**	**77.961**	**26.034**	**17.492**	**20.619**	**12.061**	40.941	**40.422**	61.524
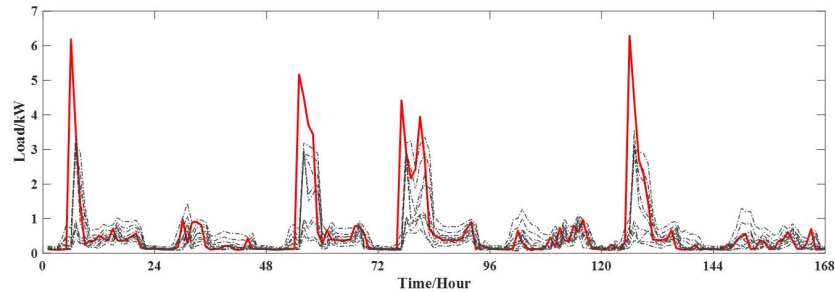
Fig. 7. Probabilistic load forecasts of consumer #1002 of one week from July 20, 2010 to July 26, 2010, where the red line is the real values and the dotted line are forecasted quantiles.

TABLE III
PINBALL LOSSES OF THE INDIVIDUAL AND COMBINATION METHODS FOR INDIVIDUAL CONSUMERS

Meter ID Methods	1002	1003	1004	1005	1008	1013	1015	1016	1017	1018
BI	0.0548	0.1726	0.2858	0.2215	0.2489	0.0772	0.1071	**0.1459**	0.1214	0.1585
NS	0.059	0.1949	0.3375	0.269	0.2873	0.0871	0.1238	0.1834	0.1351	0.1828
MED	0.0542	0.1716	0.2845	0.2236	0.2504	0.0777	0.107	0.1539	0.1219	0.1577
SA	0.054	0.1707	0.2836	0.2211	0.2491	0.0775	0.1067	0.1518	0.1215	0.1573
WA	0.054	0.1707	0.2836	0.2211	0.249	0.0775	0.1067	0.1518	0.1214	0.1573
QRA-E	0.0567	0.1737	0.3031	0.234	0.259	0.0782	0.1104	0.1562	0.1273	0.1645
QRA-A	0.0578	0.186	0.3103	0.2612	0.2664	0.0826	0.1138	0.1729	0.1285	0.1686
QRA-T	0.0548	0.1709	0.288	0.2231	0.2526	0.077	0.1081	0.1493	0.1219	0.1576
CQRA-E	0.0696	0.2157	0.3668	0.2787	0.317	0.0995	0.134	0.1824	0.1586	0.1996
CQRA-A	0.0535	0.1705	0.2876	0.2257	0.252	**0.0766**	0.1075	0.1504	**0.1202**	**0.1565**
CQRA-T	**0.0532**	**0.1689**	**0.2811**	**0.2185**	**0.2458**	0.0768	**0.1064**	0.1525	0.1208	0.1574

2) Results: Fig. 7 shows the hour-ahead probabilistic load forecasting results of consumer #1002. In contrast to the system-level load profiles, the load profile shows great volatility; thus the 10-th and 90-th quantiles do not effectively cover the spikes.

Table III illustrates the pinball losses of the different methods for the 10 selected individual consumers. It can be seen that the proposed combining method has the best performance for all consumers except consumer #1016. For consumer #1016, the sample averaging, weighted averaging, QRA, and the proposed averaging do not outperform the best individual method. Again, the results show that the ensemble method does not always have better performance than individual methods. Another interesting observation is that two QRA models have worse performance compared with the best individual for all selected individual consumers. It means that QRA model may not be able to handle high uncertainty load profile such as

the individual residential load. However, the proposed combining method is robust for this situation. The essential difference between our proposed combining method and QRA is that the proposed resemble method determines the optimal weights of the most relevant N quantiles while QRA determines the optimal weights of all the $N \times Q$ quantiles. Large number of regressors (quantiles to be regressed) in QRA contain many less relevant regressors and may lead to overfitting.

VII. CONCLUSION AND FUTURE WORK

This paper proposes a constrained quantile regression averaging method for quantile forecasting to take full advantage of the increasing number of forecasting models that have been described in the literature. The ensemble problem is formulated as a series of LP problems, each corresponding to a quantile. The results show that the proposed ensemble method

effectively improves the forecasting performance in terms of pinball loss compared with individual models.

Future works will consider two aspects. The first is the investigation of pruning methods such as lasso-based method that can be integrated into the proposed combination method refine the individual probabilistic forecasting model and further improve the performance of the combined model. The second is the extension of our method to probabilistic renewable energy forecasting. Note that our proposed method is performed on quantile results and is not limited to only load forecasting. It is interesting to investigate the improvements of the proposed method on different load datasets and renewable energy forecasting.

REFERENCES

[1] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of smart meter data analytics: Applications, methodologies, and challenges," *IEEE Trans. Smart Grid*, to be published, doi: 10.1109/TSG.2018.2818167.

[2] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *Int. J. Forecast.*, vol. 32, no. 3, pp. 914–938, 2016.

[3] Y. Wang *et al.*, "An efficient approach to power system uncertainty analysis with high-dimensional dependencies," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 2984–2994, May 2018.

[4] T. Hong, P. Pinson, and S. Fan, "Global energy forecasting competition 2012," *Int. J. Forecast.*, vol. 30, no. 2, pp. 357–363, 2014.

[5] T. Hong, P. Wang, and L. White, "Weather station selection for electric load forecasting," *Int. J. Forecast.*, vol. 31, no. 2, pp. 286–295, 2015.

[6] P. Wang, B. Liu, and T. Hong, "Electric load forecasting with recency effect: A big data approach," *Int. J. Forecast.*, vol. 32, no. 3, pp. 585–597, 2016.

[7] F. L. Quilumba, W.-J. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, "Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 911–918, Mar. 2015.

[8] Y. Wang, Q. Chen, M. Sun, C. Kang, and Q. Xia, "An ensemble forecasting method for the aggregated load with sub profiles," *IEEE Trans. Smart Grid*, to be published, doi: 10.1109/TSG.2018.2807985

[9] F. Ziel, "Modeling public holidays in load forecasting: A German case study," *J. Mod. Power Syst. Clean Energy*, vol. 6, no. 2, pp. 191–207, Mar. 2018.

[10] J. Xie and T. Hong, "Load forecasting using 24 solar terms," *J. Mod. Power Syst. Clean Energy*, vol. 6, no. 2, pp. 208–214, 2018.

[11] J. Xie, Y. Chen, T. Hong, and T. D. Laing, "Relative humidity for load forecasting models," *IEEE Trans. Smart Grid*, vol. 9, no. 1, pp. 191–198, Jan. 2018.

[12] J. Xie and T. Hong, "Wind speed for load forecasting models," *Sustainability*, vol. 9, no. 5, p. 795, 2017.

[13] J. Luo, T. Hong, and S.-C. Fang, "Benchmarking robustness of load forecasting models under data integrity attacks," *Int. J. Forecast.*, vol. 34, no. 1, pp. 89–104, 2018.

[14] J. Luo, T. Hong, and M. Yue, "Real-time anomaly detection for very short-term load forecasting," *J. Mod. Power Syst. Clean Energy*, vol. 6, no. 2, pp. 235–243, Mar. 2018.

[15] M. Cui, J. Zhang, Q. Wang, V. Krishnan, and B. M. Hodge, "A data-driven methodology for probabilistic wind power ramp forecasting," *IEEE Trans. Smart Grid*, to be published. doi: 10.1109/TSG.2017.2763827.

[16] R. J. Hyndman and S. Fan, "Density forecasting for long-term peak electricity demand," *IEEE Trans. Power Syst.*, vol. 25, no. 2, pp. 1142–1153, May 2010.

[17] T. Hong *et al.*, "Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond," *Int. J. Forecast.*, vol. 32, no. 3, pp. 896–913, 2016.

[18] P. Gaillard, Y. Goude, and R. Nedellec, "Additive models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting," *Int. J. Forecast.*, vol. 32, no. 3, pp. 1038–1050, 2016.

[19] V. Dordonnat, A. Pichavant, and A. Pierrot, "GEFCom2014 probabilistic electric load forecasting using time series and semi-parametric regression models," *Int. J. Forecast.*, vol. 32, no. 3, pp. 1005–1011, 2016.

[20] J. Xie and T. Hong, "GEFCom2014 probabilistic electric load forecasting: An integrated solution with forecast combination and residual simulation," *Int. J. Forecast.*, vol. 32, no. 3, pp. 1012–1016, 2016.

[21] J. Xie and T. Hong, "Temperature scenario generation for probabilistic load forecasting," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 1680–1687, May 2018.

[22] J. Xie, T. Hong, T. Laing, and C. Kang, "On normality assumption in residual simulation for probabilistic load forecasting," *IEEE Trans. Smart Grid*, vol. 8, no. 3, pp. 1046–1053, May 2017.

[23] B. Liu, J. Nowotarski, T. Hong, and R. Weron, "Probabilistic load forecasting via quantile regression averaging on sister forecasts," *IEEE Trans. Smart Grid*, vol. 8, no. 2, pp. 730–737, Mar. 2017.

[24] S. B. Taieb, R. Huser, R. J. Hyndman, and M. G. Genton, "Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2448–2455, Sep. 2016.

[25] D. Gan, Y. Wang, S. Yang, and C. Kang, "Embedding based quantile regression neural network for probabilistic load forecasting," *J. Mod. Power Syst. Clean Energy*, vol. 6, no. 2, pp. 244–254, 2018.

[26] J. Xie and T. Hong, "Variable selection methods for probabilistic load forecasting: Empirical evidence from seven states of the United States," *IEEE Trans. Smart Grid*, to be published, doi: 10.1109/TSG.2017.2702751.

[27] Y. Wang *et al.*, "Data-driven probabilistic net load forecasting with high penetration of behind-the-meter PV," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3255–3264, May 2018.

[28] S. Arora and J. W. Taylor, "Forecasting electricity smart meter data using conditional kernel density estimation," *Omega*, vol. 59, pp. 47–59, Mar. 2016.

[29] P. Kou and F. Gao, "A sparse heteroscedastic model for the probabilistic load forecasting in energy-intensive enterprises," *Int. J. Elect. Power Energy Syst.*, vol. 55, pp. 144–154, Feb. 2014.

[30] J. Mendes-Moreira, C. Soares, A. M. Jorge, and J. F. D. Sousa, "Ensemble approaches for regression: A survey," *ACM Comput. Surveys*, vol. 45, no. 1, p. 10, 2012.

[31] J. Nowotarski, B. Liu, R. Weron, and T. Hong, "Improving short term load forecast accuracy via combining sister forecasts," *Energy*, vol. 98, pp. 40–49, Mar. 2016.

[32] S. Li, P. Wang, and L. Goel, "A novel wavelet-based ensemble method for short-term load forecasting with hybrid neural networks and feature selection," *IEEE Trans. Power Syst.*, vol. 31, no. 3, pp. 1788–1798, May 2016.

[33] J. Xie, B. Liu, X. Lyu, T. Hong, and D. Basterfield, "Combining load forecasts from independent experts," in *Proc. North Amer. Power Symp. (NAPS)*, Charlotte, NC, USA, 2015, pp. 1–5.

[34] G. Dudek, "Heterogeneous ensembles for short-term electricity demand forecasting," in *Proc. Int. Sci. Conf. Elect. Power Eng. (EPE)*, May 2016, pp. 1–6.

[35] S. G. Hall and J. Mitchell, "Combining density forecasts," *Int. J. Forecast.*, vol. 23, no. 1, pp. 1–13, 2007.

[36] M. P. Clements and D. I. Harvey, "Combining probability forecasts," *Int. J. Forecast.*, vol. 27, no. 2, pp. 208–223, 2011.

[37] A. S. Jore, J. Mitchell, and S. P. Vahey, "Combining forecast densities from VARs with uncertain instabilities," *J. Appl. Econometrics*, vol. 25, no. 4, pp. 621–634, 2010.

[38] A. Bracale, G. Carpinelli, and P. De Falco, "A probabilistic competitive ensemble method for short-term photovoltaic power forecasting," *IEEE Trans. Sustain. Energy*, vol. 8, no. 2, pp. 551–560, Apr. 2017.

[39] E. Mangalova and O. Shesterneva, "Sequence of nonparametric models for GEFCom2014 probabilistic electric load forecasting," *Int. J. Forecast.*, vol. 32, no. 3, pp. 1023–1028, 2016.

[40] S. Haben and G. Giasemidis, "A hybrid model of kernel density estimation and quantile regression for GEFCom2014 probabilistic load forecasting," *Int. J. Forecast.*, vol. 32, no. 3, pp. 1017–1022, 2016.

[41] C. Chen, "A finite smoothing algorithm for quantile regression," *J. Comput. Graph. Stat.*, vol. 16, no. 1, pp. 136–164, 2007.

[42] J. H. Friedman, "Stochastic gradient boosting," *Comput. Stat. Data Anal.*, vol. 38, no. 4, pp. 367–378, 2002.

[43] J. Larsen and C. Goutte, "On optimal data split for generalization estimation and model selection," in *Proc. IEEE Signal Process. Soc. Workshop Neural Netw. Signal Process. IX*, Madison, WI, USA, 1999, pp. 225–234.

[44] C. W. Granger, "Combining forecasts—Twenty years later," *J. Forecast.*, vol. 8, no. 3, pp. 167–173, 1989.

[45] R. Koenker, *Quantile Regression* (Econometric Society Monographs). Cambridge, U.K.: Cambridge Univ. Press, 2005.

[46] V. Chernozhukov, I. Fernández-Val, and A. Galichon, "Quantile and probability curves without crossing," *Econometrica*, vol. 78, no. 3, pp. 1093–1125, 2010.

[47] P. J. Brown, M. Vannucci, and T. Fearn, "Bayes model averaging with selection of regressors," *J. Roy. Stat. Soc. B (Stat. Methodol.)*, vol. 64, no. 3, pp. 519–536, 2002.

[48] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky, "Bayesian model averaging: A tutorial," *Stat. Sci.*, vol. 14, no. 4, pp. 382–401, 1999.

[49] A. E. Faria and E. Mubwandarikwa, "The geometric combination of Bayesian forecasting models," *J. Forecast.*, vol. 27, no. 6, pp. 519–535, 2008.

[50] C. Conflitti, C. De Mol, and D. Giannone, "Optimal combination of survey forecasts," *Int. J. Forecast.*, vol. 31, no. 4, pp. 1096–1103, 2015.

[51] A. J. Cannon, "Quantile regression neural networks: Implementation in R and application to precipitation downscaling," *Comput. Geosci.*, vol. 37, no. 9, pp. 1277–1284, 2011.

[52] N. Meinshausen. (2016). *quantregForest: Quantile Regression Forests*. [Online]. Available: https://CRAN.R-project.org/package=quantregForest

[53] Greg Ridgeway With Contributions From Others. (2017). *GBM: Generalized Boosted Regression Models, R Package Version 2.1.3*. [Online]. Available: https://CRAN.R-project.org/package=gbm

[54] J. Lofberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *Proc. IEEE Int. Conf. Robot. Autom.*, New Orleans, LA, USA, Sep. 2004, pp. 284–289.

[55] ISO New England. (2017). *ISO New England Zonal Information*. [Online]. Available: https://www.iso-ne.com/isoexpress/web/reports/pricing/-/tree/zone-info
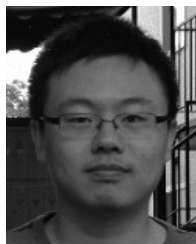
Yushi Tan received the B.S. degree from the Electrical Engineering Department, Tsinghua University, China, in 2013.

He is currently pursuing the Ph.D. degree with the University of Washington, Seattle, WA, USA. His research interests include load forecasting and power system resilience.

Tao Hong received the B.Eng. degree in automation from Tsinghua University, Beijing, China, in 2005, and the Ph.D. degree in operation research and electrical engineering from North Carolina State University, Raleigh, NC, USA, in 2010.

He is the Director of the Big Data Energy Analytics Laboratory, the Graduate Program Director, an Assistant Professor of system engineering and engineering management, and an Associate of the Energy Production and Infrastructure Center with the University of North Carolina at Charlotte, Charlotte, NC, USA. He has authored the book entitled *Electric Load Forecasting: Fundamentals and Best Practices* (OTexts, 2015), and a blog Energy Forecasting since 2013. He is the Founding Chair of the IEEE Working Group on Energy Forecasting and the General Chair of the Global Energy Forecasting Competition.

Yi Wang (S'14) received the B.S. degree from the Department of Electrical Engineering, Huazhong University of Science and Technology, Wuhan, China, in 2014.

He is currently pursuing the Ph.D. degree with Tsinghua University. He is also a Visiting Student Researcher with the University of Washington, Seattle, WA, USA. His research interests include data analytics in smart grid and multiple energy systems.

Daniel S. Krischen (M'86–SM'91–F'07) received the degree in electrical and mechanical engineering from the Universite Libre de Bruxelles, Brussels, Belgium, in 1979, and the M.Sc. and Ph.D. degrees from the University of Wisconsin, Madison, WI, USA, in 1980 and 1985, respectively.

He is currently a Close Professor of electrical engineering with the University of Washington, Seattle, WA, USA. His research interests include smart grids, the integration of renewable energy sources in the grid, power system economics, and power system security.

Ning Zhang (M'12–SM'18) received the B.S. and Ph.D. degrees from the Electrical Engineering Department, Tsinghua University, China, in 2007 and 2012, respectively.

He is currently an Associate Professor with Tsinghua University. His research interests include multiple energy systems integration, renewable energy, and power system planning and operation.

Chongqing Kang (M'01–SM'08–F'17) received the Ph.D. degree from the Department of Electrical Engineering, Tsinghua University, Beijing, China, in 1997.

He is currently a Professor with Tsinghua University. His research interests include power system planning, power system operation, renewable energy, low carbon electricity technology, and load forecasting.