

Short Term Probabilistic Load Forecasting With Integrated Methods

Manying Zhang¹, Lei Wang¹, Weimin Zheng², Hongqiao Peng³, Yue Zhu^{*3} and Jie Gu³

¹Electric Power Limited Company, Economic Research Institute, 310007 Hangzhou, China

²Electric Power Limited Company, State Grid Zhejiang, 310007 Hangzhou, China

³Shanghai Jiao Tong University, Department of Electrical Engineering, 200240 Shanghai, China

Abstract. In smart grid era, electric load is becoming more stochastic and less predictable in short horizons with more intermittent energy and competitive electricity market transactions. As a result, short-term probabilistic load forecasting (STPLF) is becoming essential for energy utilities because it helps quantify the risks of decision-making for power systems operation. Currently, probabilistic load forecasts (PLF) are commonly produced from three single components, namely input, model and output. Nevertheless, whether integrating two components to represent dual uncertainties of electric load is practical and able to improve STPLF attracts little regards. To address this issue, this paper proposes three integrated methods by pairwise combination of single representative component, i.e. uniform-biased temperature scenarios (UBTS), quantile regression (QR) and logarithmic residual empirical simulation (LRES). Case study on real utility data demonstrates the superiority of the integrated methods and excavates the relationship between predictive model class and specific integrated method.

1 Introduction

Grid modernization with more intermittent energy and competitive electricity market transactions has exacerbated the volatility and uncertainty of electric load, making it less predictable even in short horizons than ever before. As a result, probabilistic load forecasting, offering intervals, quantiles and probability density as prediction formats, has become significant and urgent for utilities to help make risk analysis and practical decision in power systems operation and planning [1]. Practical applications of probabilistic load forecasting include probabilistic load flow [2], stochastic unit commitment [3] and energy management [4].

By far, researchers and forecasters mainly produced probabilistic load forecasts from three single components of forecasting engines [5], i.e. (1) the **input** component: simulating predictors to generate input scenarios; (2) the **model** component: constructing probabilistic forecasting models or multiple predictive models; (3) the **output** component: output post-processing via residual simulation or forecasting combination.

Among various generation methods of PLF, temperature scenarios, quantile regression and residual simulation are the most representative methods for the aforementioned three components, respectively. For example, in the **input** component, Xie and Hong [6] adopted fixed-date, shifted-date and bootstrap methods to generate temperature scenarios for probabilistic load forecasting; in the **model** component, Yang *et al.* [7] proposed Gaussian process quantile regression, a dual

probabilistic model, to produce probability density for electric load; in the **output** component, Fan and Hyndman [8] generated day-ahead prediction intervals for electric load in Australian National Electricity Market by simulating forecasting residuals via block bootstrapping.

Since the probabilistic generation method from three single components of forecasting engines only reveals a specific uncertainty of electric load [9], a research question is naturally raised, “*whether combining two kinds of uncertainties via integrating pairwise components could help improve probabilistic load forecasts?*” To our best knowledge, little literature focused on this topic, which makes the integrated methods for probabilistic load forecasts a blank to industry and research.

This paper contributes to formally investigating the integrated methods for probabilistic load forecasting. We first pick out uniform-biased temperature scenarios (UBTS), quantile regression (QR) and logarithmic residual empirical simulation (LRES) as three representatives for probabilistic generation methods from three single components. Through pairwise combination of single representative, we then propose three integrated methods, a.k.a UBTS-QR, QR-LRES and LRES-UBTS. Compared with probabilistic generation method from single component, whether integrated methods help improve probabilistic load forecasts can be validated. Besides, we also excavate the relationship between a certain integrated method and predictive models.

* Corresponding author: 282130460@qq.com

The rest of this paper is organized as follows: Section II elaborates three representative probabilistic generation methods and then illustrates how to integrate the representative generation methods from two components of forecasting engines. In Section III, two model classes, i.e. Multi-linear regression (MLR) based and Gradient Boosting Regression Tree (GBRT) based models with different predictive powers are briefly introduced. Section IV presents the case studies on Shanghai utility data to evaluate the performances of integrated methods. Section V concludes the paper and provides future research directions.

2 Integrated Methods

2.1 Probabilistic generation methods from single component

Since the proposed integrated methods are derived from the probabilistic generation methods from single component, we first briefly introduce three representatives for three single components of forecasting engines, respectively.

2.1.1 Input component: uniform-biased temperature scenario (UBTS)

As temperature forecasts plays an important role in load forecasting [1], it is intuitive to generate temperature scenarios on the basis of temperature forecasts. To our best knowledge, day-ahead hourly temperatures can be precisely predicted with the maximum mean absolute error around 2 °C [10].

In this paper, we apply uniform-distributed noises onto temperature forecasts to generate day-ahead temperature scenarios, denoted as Uniform-biased temperature scenarios. According to Ref. [11], we set the temperature forecasting biases follow the uniform distribution, $[-2, 2](^{\circ}\text{C})$. Practically, we simulate the uniform biases for 99 times, then 99 day-ahead temperature scenarios and 99 point load forecasts will be generated.

2.1.2 Model component: quantile regression (QR)

Quantile regression (QR), capable of creating multiple predictive models by varying any quantiles, is a model-based probabilistic generation method. Theoretically, QR-based models keep the same structure as point forecast model, but distinguish parameters' values according to certain quantiles by minimizing the pinball loss function (Eq.(9)). In other words, m quantiles will produce m different predictive models. More details about QR can be found in [11].

In this paper, quantiles q shade from 0.01 to 0.99 with an interval of 0.01. Therefore, 99 quantile forecasts can be generated from QR method.

2.1.3 Output component: logarithmic residual empirical simulation (LRES)

Residual simulation is an effective way in output side to move from single point forecast to probabilistic forecasts. Since no statistical evidence supports forecasting residuals generally follow a certain distribution, logarithmic residual empirical simulation (LRES) is proposed to generate day-ahead forecasting residuals.

To simulate residuals, we first apply logarithmic transformation to load data, then extract the forecasting residuals in the past seven days and take 99 percentiles from empirical distribution. Next, 99 empirical-distributed residuals are added back to the point forecast followed by the exponent transformation to generate 99 quantile forecasts in the forecasting horizon.

2.2. Integrated probabilistic generation methods

After introducing three representative probabilistic generation methods from single component of forecasting engine, three kinds of integrated methods (see Table 1) are proposed by pairwise combination of the aforementioned representatives.

Table 1. Integrated generation methods by pairwise combination

Dual Components	Integrated Generation Method
Input and Model	UBTS-QR
Output and Input	LRES-UBTS
Model and Output	QR-LRES

2.2.1 UBTS-QR

The combination of temperature scenarios and quantile regression takes both the uncertainties of temperature forecasts and load variation into account. The core of this integration is to feed the uniform-biased temperature scenarios into QR based predictive models.

In details, we first train multiple predictive models following the QR method, then replace the temperature forecast (usually a deterministic value given) with uniform-biased temperature scenarios in the forecasting horizon and take these temperature scenarios as predictors for the trained QR based models. In other words, if m QR based models are trained and J temperature scenarios are generated, then $m \times J$ point forecasts will be produced. Finally, 99 quantile forecasts are derived from the empirical distribution of $m \times J$ point forecasts.

2.2.2 LRES-UBTS

The association between temperature scenarios with residual simulation originates from the mismatches of temperature forecasts and predictive models. Integrating

uniform bias temperature scenarios and logarithmic residual empirical simulation helps make correction for temperature-scenario-based point forecasts.

Specifically, residuals over past seven days are first determined for each temperature scenario based forecast, then their 99 percentiles are extracted from the empirical distribution and added back to the corresponding temperature-scenario-based point forecast. Note that all the residuals are logarithmic and the transformation should be accordingly performed.

2.2.3 QR-LRES

Integrating quantile regression and residual simulation takes dual uncertainties of load and predictive models into consideration. This approach attempts to correct the mismatches of the certain q th-quantile predictive model.

Similar to UBTS-LRES, we first calculate the logarithmic residuals over past seven days for each QR based model, then take their 99 percentiles from the empirical distribution and adds them back to the corresponding quantile forecast as well as applying the exponent transformation.

3 Predictive models

Probabilistic generation methods convert point load forecasts into probabilistic forecasts. Before applying and evaluating these methods, we need underlying predictive models to produce point forecasts. To ensure the generality of our findings, Multi-Linear Regression (MLR) and Gradient Boosting Regression Tree (GBRT) model classes, with different variables input to reflect different levels of predictive powers, will be constructed.

3.1 Multi-Linear regression class

The first model in MLR class, named as *Vanilla*, was first proposed by Hong [12] and served as benchmarking STLF model in GEFCom2012 [13]. *Vanilla*, mainly featured as interacting temperature and calendar variables, can be expressed as:

$$\hat{y}_t = \beta_0 + \beta_1 Trend_t + \beta_2 M_t + \beta_3 D_t + \beta_4 H_t + \beta_5 D_t H_t + f(T_t) \quad (1)$$

Where, \hat{y}_t represents hourly load forecasts at time t , β_i are parameters to be estimated, and T is the temperature at time t . Besides, $Trend_t$ indicates the linear trend at time t and its value increases in natural order as t moves forward. M_t , D_t , H_t are class variables indicating month-of-the-year, day-of-the-week, and hour-of-the-day, respectively; and

$$f(T_t) = \beta_6 T_t + \beta_7 T_t^2 + \beta_8 T_t^3 + \beta_9 T_t M_t + \beta_{10} T_t^2 M_t + \beta_{11} T_t^3 M_t + \beta_{12} T_t H_t + \beta_{13} T_t^2 H_t + \beta_{14} T_t^3 H_t \quad (2)$$

More details for *Vanilla* can be found in [12].

The second model, removing the calendar variables but adds the recent loads as predictors, is denoted as *Lag-Cubic*. In STLF, lagged loads are commonly used as predictors [8], [12] for their ability to capture serial correlations. Since model construction is not the point of our research, the lagged loads are heuristically selected.

Here, the lagged loads for 1 day, 2 days and 1 week are selected. It can be written as:

$$\hat{y}_t = \beta_0 + \beta_1 Trend_t + \beta_2 T_t + \beta_3 T_t^2 + \beta_4 T_t^3 + \beta_5 y_{t-24} + \beta_6 y_{t-48} + \beta_7 y_{t-168} \quad (3)$$

The third model, containing calendar variables, temperatures and lagged loads as predictors, is extended by *Vanilla*. Within it, y_{t-24} is attached and this model is denoted as *E-Vanilla*:

$$\hat{y}_t = \beta_0 + \beta_1 Trend_t + \beta_2 M_t + \beta_3 D_t + \beta_4 H_t + \beta_5 D_t H_t + f(T_t) + \beta_{15} y_{t-24} \quad (4)$$

3.2 Gradient Boosting Regression Tree Class

GBRT, a decision-tree-based ensemble regression model, is trained by adding a new classification and regression tree (CART) in an iterative process. The optimal CART for k th iteration is determined by fitting residuals from the $(k-1)$ th tree. Given a specific loss function $L(y_t, f(\mathbf{X}_t))$, the residuals for time t at k th iteration can be written as:

$$\gamma_{k,t} = - \left. \frac{\partial L(y_t, f(\mathbf{X}_t))}{\partial f(\mathbf{X}_t)} \right|_{f(\mathbf{X}_t) = f_{k-1}(\mathbf{X}_t)} \quad (5)$$

K is denoted as the number of iterations and N_{GB} is denoted as the maximum number of terminal nodes of each learned tree and λ the learning rate. The new CART at k th iteration can fitted by $(\mathbf{X}_t, \gamma_{k,t})$ with N_{GB} disjoint regions $\{R_{k,n_{GB}}\}_1^{N_{GB}}$. The optimal value for each terminal node (denoted as $\gamma_{k,n_{GB}}$) can be further advanced by solving:

$$\gamma_{k,n_{GB}} = \arg \min_{\gamma} \sum_{\mathbf{X}_t \in R_{k,n_{GB}}} L(y_t, f_{k-1}(\mathbf{X}_t) + \gamma) \quad (6)$$

Next, the gradient boosting tree at k th iteration can be calculated:

$$f_k(\mathbf{X}_t) = f_{k-1}(\mathbf{X}_t) + \lambda \sum_{n_{GB}=1}^{N_{GB}} \gamma_{k,n_{GB}} \mathbf{I}(\mathbf{X}_t \in R_{k,n_{GB}}) \quad (7)$$

After K iterations, the GBRT is finally determined:

$$\hat{y}_t = f(\mathbf{X}_t) = f_0(\mathbf{X}_t) + \lambda \sum_{k=1}^K \sum_{n_{GB}=1}^{N_{GB}} \gamma_{k,n_{GB}} \mathbf{I}(\mathbf{X}_t \in R_{k,n_{GB}}) \quad (8)$$

Where, $f_0(\mathbf{X}_t)$ denotes the initial CART, a weak learned tree, trained by data set $\mathbf{D} = \{(\mathbf{X}_t, y_t)\}_1^T$. More details about GBRT can be found in [14].

With the framework of GBRT, three GBRT-based models can be constructed by feeding different variables. Without losing justifiability, the same variables as MLR class are selected for benchmarking GBRT (denoted as *B-GBRT*), lagged GBRT (denoted as *Lag-GBRT*) and extended GBRT (denoted as *E-GBRT*), respectively. In details, the input variables are listed in Table 2:

Table 2. Input variables for gradient boosting regression tree class

Models	Input variables
<i>B-GBRT</i>	M_t, D_t, H_t, T_t

<i>Lag-GBRT</i>	$T_t, y_{t-24}, y_{t-48}, y_{t-168}$
<i>E-GBRT</i>	$M_t, D_t, H_t, T_t, y_{t-24}$

4 Results and discussion

4.1 Data source

Shanghai, one of the most developed cities around the world, is located in East China. Her utility provides energy and related services for 15 districts and 1 county with around 25 million people. In this paper, hourly electric load and hourly temperature information from 2013 to 2016 in Shanghai, part of the data used in [15], are utilized.

Note that the first three years' data (2013-2015) are used for training the model parameters, while the last year's (2016) data intended for short term ex ante probabilistic load forecasting. The forecasting horizon is fixed at one day ahead.

4.2 Evaluation metric

To make fair evaluation for the proposed integrated methods on probabilistic load forecasts, quantile score is used to evaluate the performances of probabilistic forecasts. It is derived from the pinball loss function, which is considered as an error measurement formally carried out in the GEFCom2014.[5] For q th quantile, if the quantile forecast is $\hat{y}_{t,q}$, then the pinball loss function can be expressed as:

$$Pinball(\hat{y}_{t,q}, y_{t,q}) = \begin{cases} (1-q)(\hat{y}_{t,q} - y_t) & y_t < \hat{y}_{t,q} \\ q(y_t - \hat{y}_{t,q}) & y_t \geq \hat{y}_{t,q} \end{cases} \quad (9)$$

To obtain quantile score, the pinball losses are summed for all targeted quantiles and then the sum is averaged over all quantiles and samples. The quantile score can be specified as:

$$QS = \frac{1}{NM} \sum_{t=1}^N \sum_{m=1}^M Pinball(\hat{y}_{t,q}, y_{t,q}) \quad (10)$$

Where, N represents the total samples of the holdout set, and M is the total number of quantiles.

4.3 Probabilistic load forecasts with MLR class

Under the MLR model class, Table 3 presents the quantile scores of the test year (2016) via applying various probabilistic generation methods from single or dual components of forecasting engine. The lowest quantile score for each predictive model is highlighted in bold. We can clearly obtain that QR-LRES method provides the best short term probabilistic load forecasts for the above MLR models.

Table 3. Input variables for gradient boosting regression tree class

Models	<i>Vanilla</i>	<i>Lag-Cubic</i>	<i>E-Vanilla</i>
--------	----------------	------------------	------------------

UBTS	423.14	401.36	241.97
QR	371.09	304.04	205.62
LRES	319.95	306.95	208.90
UBTD-QS	368.00	303.96	204.17
QR-LRES	304.08	303.53	197.04
LRES-UBTS	320.57	306.56	207.88

To make the comparison between dual-component integrated methods and single-component generation methods more intuitive, a color chart is drawn to indicate the degree of relative improvement as to better and worse method for integration, as Fig. 1 shows. It can be seen that QR-LRES outperforms the better method most, which might account for the fact that QR-LRES matching MLR model class can improve probabilistic load forecasts to the largest extent. An evident of why QR-LRES can beat the better method (QR) in quantile score is shown in Fig. 2, where actual loads in summer and winter peak week are more central within the intervals created by QR-LRES than those by QR based on *E-Vanilla*.

Relative Improvement Compared to Better method from Single Component				
Models Methods	<i>Vanilla</i>	<i>Lag-Cubic</i>	<i>E-Vanilla</i>	Avg.
UBTS-QR	0.83%	0.03%	0.71%	0.52%
QR-LRES	4.96%	0.17%	4.17%	3.10%
LRES-UBTS	-0.19%	0.13%	0.49%	0.14%
Relative Improvement Compared to Worse method from Single Component				
Models Methods	<i>Vanilla</i>	<i>Lag-Cubic</i>	<i>E-Vanilla</i>	Avg.
UBTS-QR	13.03%	24.27%	15.62%	17.64%
QR-LRES	18.06%	1.11%	5.68%	8.28%
LRES-UBTS	24.24%	23.62%	14.09%	20.65%

Fig. 1. Relative improvements of integrated methods on MLR class.

In addition, the integrated methods override 8 out of 9 on quantile scores compared to the better single-component method and outperform all compared to the worse one. Therefore, we can conclude that *integrated methods by combining pairwise components help improve short term probabilistic load forecasts.*

4.4 Discussion on GBRT class

To avoid drawing the conclusion specifically for one model class, we conduct the same short term ex ante probabilistic load forecasting with GBRT model class. Table 4 shows the quantile scores of the test year (2016) with the same generation methods but based on GBRT-based models. The values within the parenthesis are the relative improvement of integrated methods as to the better single-component (left) and worse method (right).

In GBRT model class, LRES-UBTS generates the best probabilistic load forecasts on average and outperforms the better method most, which is similar to QR-LRES in MLR case. Meanwhile, the integrated methods outweigh 13 out of 18 comparisons to single-component. Again, the conclusion that *integrated methods by combining pairwise components help*

improve short term probabilistic load forecasts is empirically verified. Besides, since the optimal integrated method differs as the model class changes, but stays consistent within the same model class, it might originate from the fact that the performance of a specific integrated method rely heavily on the model class.

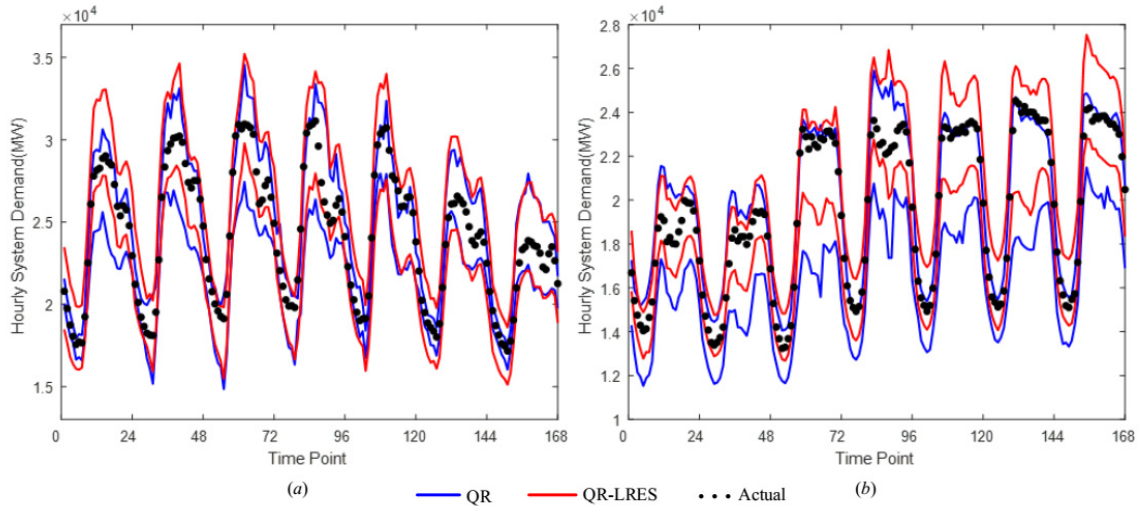


Fig. 2. Actual load (black dots) and 98% PIs (blue: QR; red: QR-LRES) in the summer (Sub-Fig.(a)) and winter (Sub-Fig.(b)) peak week. Actual loads are more central within the 98% interval of QR-LRES than QR.

Table 4. Quantile scores of the test Year (2016) with various generation methods based on GBRT class.

Models	<i>B-GBRT</i>	<i>Lag-GBRT</i>	<i>E-GBRT</i>
UBTS	565.68	380.75	290.61
QR	487.90	295.37	244.01
LRES	315.74	292.50	227.58
UBTD-QS	489.49(-0.33% 13.47%)	295.57(-0.07 22.37%)	244.19(-0.16% 15.97%)
QR-LRES	312.45(1.04% 35.96%)	294.81(-0.44% 0.19%)	236.05(-3.72% 3.26%)
LRES-UBTS	313.37(0.75% 44.60%)	292.98(0.18% 23.05%)	226.58(0.44% 22.03%)

5 Conclusion and future direction

This paper formally investigates the integrated methods for short term probabilistic load forecasting by proposing three integrated methods, a.k.a. UBTS-QR, QR-LRES and LRES-UBTS, by pairwise combining the components from the forecasting engine. The Shanghai utility case shows that integrated methods help improve short term probabilistic load forecasts. Meanwhile, the model class is the most decisive factor for the performances of a specific integrated methods. Therefore, it is practically necessary to pick out the optimal integrated method for a certain model class, which

requires numerous case studies. Following a similar framework of our proposed methods, more case studies are encouraged to conduct for other regions, model classes and forecasting horizons.

Future works will take the following two aspects into account. First, the intrinsic theory for integrating two methods should be more specific and clear, as the integrated methods only outperform the better single-component methods slightly, but greatly outweighs the worse one. Hence, there must be some trade-off between the two single-component methods for integration, which needs more investigation. Second, more density or quantile combination theory should be dug out, as the core of integration is the mixture of two distributions.

This motivates researchers to propose other ways of integration.

The authors would like to thank Shanghai Utility to provide the hourly load data for our experiment.

References

1. T. Hong, J. Wilson, J. Xie, IEEE Tran. on Smart Grid, **5**, 456–462 (2014)
2. M. Kabir, Y. Mishra, R. Bansal, Appl. Ener., **163**, (2016)
3. M. E. Lebotsa, C. Sigauke, A. Bere, R. Fildes, J. E. Boylan, Appl. Ener., **222**, 104–118 (2018)
4. L. Zhou, F. Li, X. Tong, IEEE Tran. on Smart Grid, **8**, 2694–2701 (2017)
5. T. Hong, S. Fan, Int. J. of Forec., **32**, 914–938 (2016)
6. J. Xie, T. Hong, IEEE Tran. on Smart Grid, **9**, 1680 – 1687 (2018)
7. Y. Yang, S. Li, W. Li, M. Qu, Appl. Ener., **213** (2018)
8. S. Fan, R. J. Hyndman, IEEE Tran. on Power Sys., **27**, 134–141 (2012)
9. G. Dahua, W. Yi, Y. Shuo, K. Chongqing, J. of Mod. Power Sys. and Clean Ener., **6**, 244–254 (2018)
10. S. Papantoniou, D.-D. Kolokotsa, Ener. and Build., **114**, (2016)
11. R. Koenker, G. Bassett Jr, J of the Econ. Soc., 33–50 (1978)
12. T. Hong, *Short term electric load forecasting*.(2010)
13. T. Hong, P. Pinson, S. Fan, Int. J. of Forec., **30**, 357–363 (2014)
14. J. H. Friedman, Comp. Sta. & Data Ana., **38**, 367–378 (2002)
15. H. Peng, C. Liu, K. Bai, J. Gu, PESGM IEEE, 1–5 (2018)