GEFCom2014 probabilistic electric load forecasting: An integrated solution with forecast combination and residual simulation

Article *in* International Journal of Forecasting · December 2015

DOI: 10.1016/j.ijforecast.2015.11.005

Jingrui Xie
University of North Carolina at Charlotte

ARTICLE IN PRESS
International Journal of Forecasting ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

International Journal of Forecasting

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

GEFCom2014 probabilistic electric load forecasting: An integrated solution with forecast combination and residual simulation

Jingrui Xie *, Tao Hong

University of North Carolina at Charlotte, Charlotte, NC, USA

ARTICLE INFO

Keywords:
Probabilistic forecasting
Load forecasting
Residual simulation
Regression analysis
Time series modeling
Neural networks
Forecast combination

ABSTRACT

We present an integrated solution for probabilistic load forecasting. The proposed solution was the basis for Jingrui Xie's submission to the probabilistic load forecasting track of the Global Energy Forecasting Competition 2014 (GEFCom2014), and consists of three components: pre-processing, forecasting, and post-processing. The pre-processing component includes data cleansing and temperature station selection. The forecasting component involves the development of point forecasting models, forecast combination, and temperature scenario based probabilistic forecasting. The post-processing component embodies residual simulation for probabilistic forecasting. In addition, we also discuss several other variations that were implemented during the competition.

1. Introduction

Utilities have been using point forecasts for system and financial planning for several decades. However, the worldwide modernization of the power grid has meant that the electricity demand has become more versatile and less predictable than ever before. As a result, probabilistic load forecasts have become increasingly important in helping utilities to quantify the uncertainties in the electricity demand. A recent tutorial review on probabilistic load forecasting (Hong & Fan, this issue) summarized the significant developments in this field and introduced a range of techniques for producing and evaluating probabilistic load forecasts.

In an attempt to address and overcome the challenges of probabilistic energy forecasting, the IEEE Working Group on Energy Forecasting organized the Global Forecasting Competition 2014 (GEFCom2014) (Hong et al., this issue). The probabilistic load forecasting track of GEFCom2014 provided the competitors with six years (2005–2010) of hourly load data and 10 years (2001–2010) of hourly temperature data from 25 anonymous weather stations. During each of the 15 weeks of the competition, each participating team was asked to provide a one-month-ahead probabilistic load forecast in the form of quantiles. The actual load and temperature data from the previously forecasted month was released incrementally every week. The detailed competition rules, a description of the data and a summary of the methods of selected teams are provided by Hong et al. (this issue). This paper presents Jingrui Xie's methodology in detail.

The methods and models implemented by Jingrui Xie evolved over the course of the competition. In this paper, we focus on the core solution framework, which was the basis of Jingrui Xie's implementations during GEFCom2014. As Fig. 1 shows, the solution framework consists of three components:

- The *pre-processing* component, which includes two parts. The first part uses Tao's vanilla benchmark model

* Corresponding author.
E-mail addresses: jingrui.rain.xie@gmail.com (J. Xie), hongtao01@gmail.com (T. Hong).

http://dx.doi.org/10.1016/j.ijforecast.2015.11.005

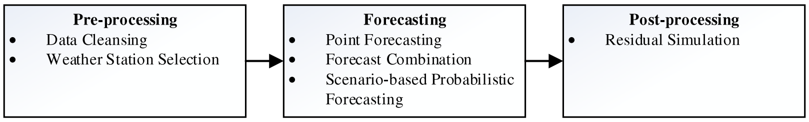

Fig. 1. Framework of Jingrui Xie's solution in GEFCom2014.

(Hong, Pinson, & Fan, 2014; Hong, Wang, & Willis, 2011) for outlier detection and data cleansing. The second part uses the same model for weather station selection, following the methodology proposed by Hong, Wang, and White (2015).

- The *forecasting* component consists of three steps. The first step sharpens the underlying point forecasting models; the second step combines the point forecasts; and the third step generates temperature scenario based probabilistic load forecasts based on the point forecasting model and 10 years of temperature history.
- The *post-processing* component simulates the residuals of the selected point forecasting models from the forecasting component, in order to improve the probabilistic forecast further.

The next three sections of this paper will discuss these three components. In Section 5, we will discuss alternative implementations and the results. The paper will then be concluded in Section 6.

2. Pre-processing

2.1. Data cleansing

The data cleansing process starts with the benchmark model in Eq. (1). It is a multiple linear regression (MLR) model with the following main and cross effects:

i. main effects: a chronological trend variable (*Trend*), the 1st to 3rd order polynomials of the temperature (T_t, T_t^2 and T_t^3), and the class variables *Month*, *Weekday* and *Hour*;
ii. cross effects: $Hour_t * Weekday_t$, $T_t * Month_t$, $T_t^2 * Month_t$, $T_t^3 * Month_t$, $T_t * Hour_t$, $T_t^2 * Hour_t$, and $T_t^3 * Hour_t$.

After estimating the benchmark model using all of the historical data, we calculate the absolute percentage error (APE) for each hourly load observation. The observations with APE values of greater than 50% are treated as outliers. For these outliers, we then replace the original observations with the predicted values from the benchmark model. This data cleansing process modifies about 0.05% of the historical load data. As an example period, Fig. 2 shows October 2–3, 2006, which this method detects as a series of outliers.

$$\begin{aligned}E(Load_t) = {} & \beta_0 + \beta_1 * Trend_t + \beta_2 * T_t + \beta_3 * T_t^2 \\ & + \beta_4 * T_t^3 + \beta_5 * Month_t + \beta_6 * Weekday_t \\ & + \beta_7 * Hour_t + \beta_8 * Hour_t * Weekday_t \\ & + \beta_9 * T_t * Month_t + \beta_{10} * T_t^2 * Month_t \\ & + \beta_{11} * T_t^3 * Month_t + \beta_{12} * T_t * Hour_t \\ & + \beta_{13} * T_t^2 * Hour_t + \beta_{14} * T_t^3 * Hour_t. \qquad (1)\end{aligned}$$

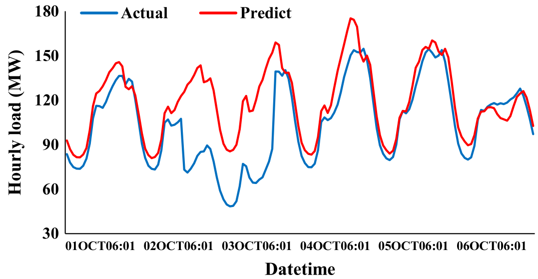

Fig. 2. Hourly actual and predicted loads for Oct. 1–6, 2006.

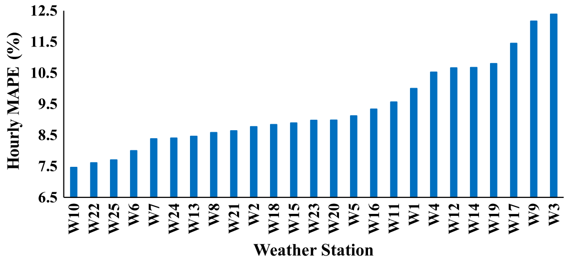

Fig. 3. Weather stations ranked by ascending MAPE (%).

2.2. Weather station selection

Hourly temperatures from 25 anonymous weather stations (W_1–W_{25}) were provided in the competition, but the competition organizer did not release any geographic information for the individual weather stations. Thus, we select the appropriate weather stations for the utility by following the weather station selection method proposed by Hong et al. (2015).

We begin by feeding the benchmark model with the temperature series from the 25 weather stations. Estimating the model using the data from the years 2007 to 2009, we obtain 25 sets of in-sample fit results, one from each of the 25 weather stations. We then calculate the mean absolute percentage errors (MAPE) of the in-sample fit results, and sort the weather stations by MAPE values in ascending order, as Fig. 3 shows.

By averaging the hourly temperature series of the top n weather stations, we can obtain combined weather stations, denoted by CW_n. For example, CW_3 is the hourly temperature series created by averaging the hourly temperatures of W_{10}, W_{22} and W_{25}. We then feed the benchmark model with each of the 25 combined temperature series in turn. Using the data from the years 2007–2009 to estimate the model, we can then forecast the year 2010. Fig. 4 shows the MAPE values of the combined weather stations for the year 2010. The best accuracy occurs for CW_{11} (i.e., the average temperature of the top 11 weather stations), which yields the lowest MAPE value of 6.18%.

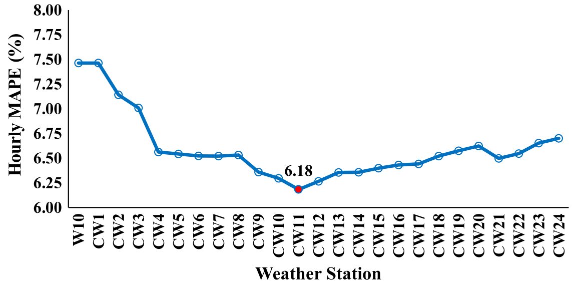

Fig. 4. MAPE (%) of combined weather stations.

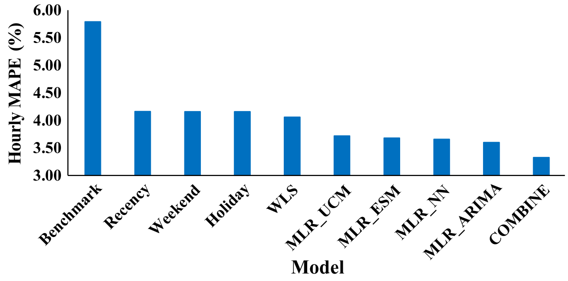

Fig. 5. Hourly MAPE (%) of the underlying point forecast models.

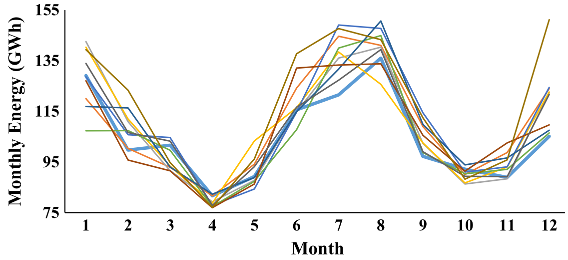

Fig. 6. Scenario-based monthly energy forecasts (year $=$ 2011).

3. Point and probabilistic forecasting

3.1. Point forecasting models

We develop point forecasting models using a two-stage approach. In the first stage, we follow the model selection process proposed by Hong (2010). The process starts with the benchmark model, except on the first day of the month. For the first day of the month, the benchmark model is expanded to include the load with a 24-hour lag, as shown in Eq. (2). It then includes special effects one by one, such as the recency effect, a weekend effect, and a holiday effect, with each special effect built upon its predecessors. Depending on the performance of these effects on the validation data set, different combinations of the aforementioned modeling effects were selected for different validation data sets in each round of the competition. At the end of the first stage, the selected model is estimated using the weighted least squares (WLS) method. The final MLR model in the first stage is used to generate a point load forecast and the residuals that are fed to the second stage.

In the second stage, we model the first-stage residuals and generate the residual forecast using four different techniques, namely unobserved component models (UCM), exponential smoothing models (ESM), three-layer feedforward artificial neural networks (ANN), and autoregressive integrated moving average models (ARIMA). We can then obtain four second-stage load forecasts by adding the four second-stage residual forecasts to the first-stage load forecast:

$$\begin{aligned} E(Load_t) = {} & \beta_0 + \beta_1 * Trend_t + \beta_2 * T_t + \beta_3 * T_t^2 \\ & + \beta_4 * T_t^3 + \beta_5 * Month_t + \beta_6 * Weekday_t \\ & + \beta_7 * Hour_t + \beta_8 * Hour_t * Weekday_t \\ & + \beta_9 * T_t * Month_t + \beta_{10} * T_t^2 * Month_t \\ & + \beta_{11} * T_t^3 * Month_t + \beta_{12} * T_t * Hour_t \\ & + \beta_{13} * T_t^2 * Hour_t + \beta_{14} * T_t^3 * Hour_t \\ & + \beta_{15} Load_{t-24}. \end{aligned} \tag{2}$$

3.2. Combining two-stage point forecasts

Combining forecasts is a practical way of enhancing the forecast accuracy. Of the various forecast combination techniques, the simple average is the easiest to implement, but still yields decent forecast improvements in many cases. Recently, a top entry of the NPower Forecasting Challenge used the simple average to combine forecasts from multiple techniques (Xie et al. 2015).

Here, we take the simple average of the four second-stage forecasts to obtain the point forecast combination. Fig. 5 shows the error statistics of these 10 point forecasting models (five from the first stage, four from the second stage, and one from forecast combination). The forecast combination yields the best accuracy, and therefore will be promoted as the underlying model for generating probabilistic forecasts. The recency model, the weekend model and the holiday model provide similar results, because only a few days, e.g., some holidays and their surrounding days, in the validation data set were significantly adjusted from the recency model due to the weekend effect and the holiday effect.

3.3. Scenario-based probabilistic load forecasting

Following the methodology proposed by Hong, Wilson, and Xie (2014), we create 10 temperature scenarios using 10 years of historical temperature data from 2001 to 2010, and use them to generate 10 load forecasts. As a result, 10 forecasts are generated for each hour based on the 10 temperature scenarios. Fig. 6 shows the monthly energy forecasts for 2011. Each line in the figure is from one of the 10 temperature scenarios. Taking the 1st–99th percentiles from the 10 point forecasts for each of the forecasted hours, we can obtain the forecasted quantiles.

4. Post-processing

As was discussed by Xie et al. (2015), residual simulation based on a normal distribution may result in some improvement relative to the probabilistic forecasts. We

Please cite this article in press as: Xie, J., & Hong, T., GEFCom2014 probabilistic electric load forecasting: An integrated solution with forecast combination and residual simulation. International Journal of Forecasting (2015), http://dx.doi.org/10.1016/j.ijforecast.2015.11.005

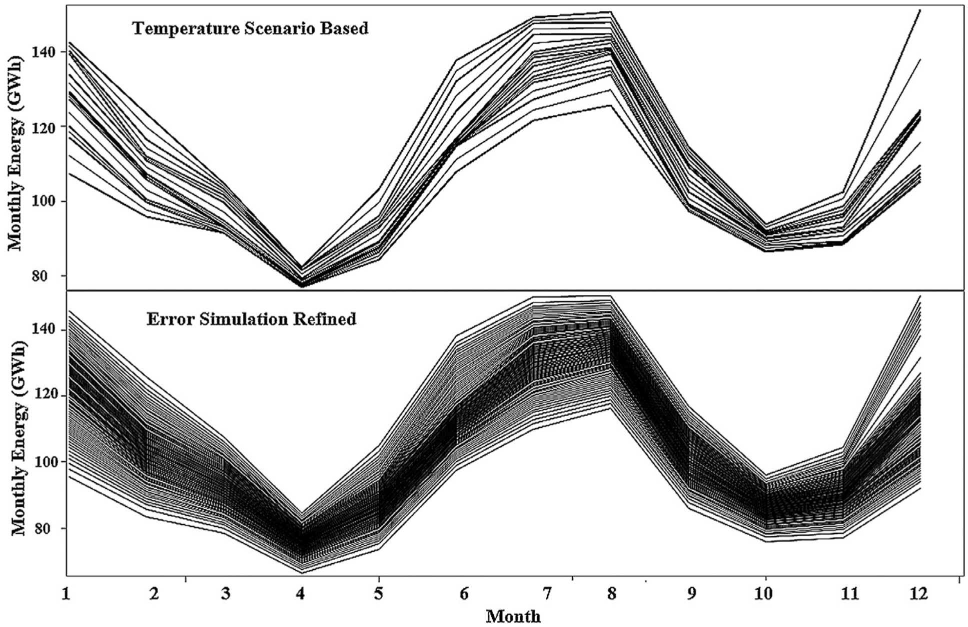

Fig. 7. Probabilistic monthly energy forecast in quantiles (year = 2011).

therefore refine the probabilistic forecasts obtained in Section 3 by modeling and simulating the residuals from the forecast combination. Note that these residuals are not the ones from the first-stage point load forecast.

We group the residuals from the same month of the previous year into 24 groups, one for each hour of the day. For each residual group, 1000 random numbers are generated from a normal distribution using the parameters derived from the empirical distribution of the residuals. The 1000 random numbers for each group are used as the simulated residuals and added back, by hour, to each probabilistic forecast generated in Section 3. A total of 10,000 forecasts are generated for each hour of the forecasted period. Finally, the 1st–99th percentiles of the 10,000 forecast results for each hour are extracted as the quantile forecasts. Fig. 7 shows two monthly energy forecasts in the form of quantiles. The one on the left is from the temperature scenario based probabilistic forecasting discussed in Section 3, while the one on the right is from the residual simulation in the post-processing step.

5. Discussion

5.1. Temperature scenario simulation

In Jingrui Xie's first trial submission (Task 1), the temperature data were shifted forward and backward to generate additional temperature scenarios. In other words, the nine years (2001–2009) of temperature data were shifted by one, two or three days to generate 27, 45 or 63 temperature scenarios, respectively, for forecasting the load in October 2010. The resulting quantile scores were 4.14, 4.14, and 4.53, respectively, while the quantile score from the original forecast based on nine scenarios is 4.09. In other words, shifting the temperature history to generate additional scenarios did not improve the original probabilistic forecast in this first trial submission. Therefore, to keep the process simple, the shifted temperature scenarios were not used to forecast the remaining months. Nevertheless, a comprehensive study is needed to determine whether shifting historical temperature data is effective or not.

5.2. More about residual simulation

Xie et al. (2015) discussed several residual grouping methods. In addition to that discussed in Section 4, we can also simulate the residuals using the same month of the previous years as one group, using weekdays of the previous years as seven groups, or using weekdays of the same month of the previous years as seven groups. Several tests showed that grouping the residuals by hour of the same month was better than the alternative methods, according to the quantile score used in GEFCom2014. Therefore, grouping residuals by hour of the same month was implemented for the rest of the competition.

The length of the residuals being used for simulation has evolved from task to task. Initially, only the residuals of the same month from the previous year were used. After Task 7, an investigation was conducted to test the impact of the residual length. The results showed that, on average, using three years of residuals yielded the best probabilistic forecasts.

5.3. Selecting point load forecasting models

Another question during the competition related to the selection of the underlying point load forecasting model. In the solution presented earlier, MAPE is used as the criterion for selecting the underlying point load forecasting model. Alternatively, we can also use the quantile score of

Please cite this article in press as: Xie, J., & Hong, T., GEFCom2014 probabilistic electric load forecasting: An integrated solution with forecast combination and residual simulation. International Journal of Forecasting (2015), http://dx.doi.org/10.1016/j.ijforecast.2015.11.005

Table 1
Quantile score for probabilistic forecasts (Year = 2011).
Bold Numbers are the Best in Each Group. Underlined Numbers are the Best across All.

	Group I (temperature scenario based)						Group II (post-processed with residual simulation)						
	Benchmark	Recency	Weekend	Holiday	WLS	COMBINED	Benchmark	Recency	Weekend	Holiday	WLS	COMBINED	Submission
1	**12.001**	12.102	12.100	12.104	12.146	12.146	**11.659**	11.830	11.831	11.831	11.959	11.962	11.867
2	10.953	10.800	10.801	**10.790**	11.008	11.008	10.621	10.560	10.561	**10.546**	10.716	10.864	10.925
3	8.607	8.566	**8.564**	8.573	8.673	8.673	**8.336**	8.437	8.434	8.442	8.526	8.534	8.438
4	5.061	4.994	4.995	4.993	4.949	**4.949**	4.878	5.008	5.007	5.005	4.895	**4.860**	4.961
5	7.580	7.341	7.337	7.317	7.283	**7.283**	7.288	7.233	7.230	7.215	**7.090**	7.173	7.275
6	**6.536**	6.847	6.847	6.845	7.173	7.173	**6.270**	6.726	6.723	6.722	6.851	7.077	6.992
7	9.437	**9.402**	9.403	9.405	9.513	9.513	9.235	9.315	9.316	9.319	**9.201**	9.494	9.052
8	11.248	**11.169**	11.170	11.173	11.194	11.194	11.109	11.122	11.123	11.128	11.090	**11.083**	11.260
9	5.898	**5.760**	5.762	5.791	5.807	5.807	5.672	**5.564**	5.568	5.594	5.646	5.729	5.486
10	4.068	**3.740**	3.743	3.782	3.833	3.833	3.542	**3.415**	3.417	3.438	3.581	3.688	3.360
11	6.505	6.253	6.255	6.255	6.158	**6.158**	6.123	6.092	6.092	**6.086**	6.0896	6.1001	5.901
12	10.818	**10.402**	10.403	10.404	10.495	10.495	10.166	9.838	9.839	**9.837**	10.171	10.277	9.732
AVG.	8.226	**8.1146**	8.1150	8.119	8.186	8.183	**7.908**	7.928	7.928	7.930	7.985	8.064	7.937

the probabilistic forecast as the measure for selecting the underlying models. In other words, we can first enumerate all of the point load forecasting models in order to generate probabilistic forecasts, then select the model with the lowest quantile score as the underlying model. However, this process was too time-consuming for Jingrui Xie to implement, as she was on maternity leave during the competition.

5.4. Comparison

Jingrui Xie's GEFCom2014 implementation is a combination of the solution introduced in Sections 3 and 4 and the alternatives discussed in Section 5. Table 1 shows the quantile scores of Jingrui Xie's GEFCom2014 submitted forecasts and 12 other forecasts. The 12 forecasts are in two groups, one with residual simulation and one without. In each group, there are six underlying models, including five from first-stage models and one from the forecast combination. On average, residual simulation in the post-processing step helps to improve the original probabilistic forecasts.

6. Conclusion

We have presented a winning solution to the probabilistic load forecasting track of GEFCom2014. In addition to the core methodology, we have also discussed several alternative implementations. The winning solution is the result of several modeling efforts, including data cleansing, weather station selection, improvements in the underlying point forecasting model, and residual simulation. While this paper focuses on describing Jingrui Xie's solution to GEFCom2014, we would also like to recognize a few areas that deserve further investigation, such as temperature scenario generation and model selection based on quantile scores.

Acknowledgments

Tao Hong, the second author of this paper, contributed significantly to the writing of this paper, but was not part of the single-person team Jingrui Xie in GEFCom2014. Dr. Shu Fan's edits and comments on the manuscript have helped improve the clarity of presentation.

References

Hong, T. (2010). *Short term electric load forecasting*. (Ph.D. dissertaion), Graduate Program of Operation Research and Dept. of Electrical and Computer Engineering, North Carolina State University.

Hong, T., & Fan, S. (2016). Probabilistic electric load forecasting: a tutorial review. *International Journal of Forecasting*, this issue.

Hong, T., Pierre, P., Fan, S., Zareipour, H., Troccoli, A., & Hyndman, R. J. (2016). Probabilistic energy forecasting: state-of-the-art. *International Journal of Forecasting*, this issue.

Hong, T., Pinson, P., & Fan, S. (2014). Global energy forecasting competition 2012. *International Journal of Forecasting*, *30*(2), 357–363.

Hong, T., Wang, P., & White, L. (2015). Weather station selection for electric load forecasting. *International Journal of Forecasting*, *31*(2), 286–295.

Hong, T., Wang, P., & Willis, H. (2011). A naïve multiple linear regression benchmark for short term load forecasting. In *2011 IEEE power and energy society general meeting, Detroit, MI, July 24–29, 2011*.

Hong, T., Wilson, J., & Xie, J. (2014). Long term probabilistic load forecasting and normalization with hourly information. *IEEE Transactions on Smart Grid*, *5*(1), 456–462.

Xie, J., Hong, T., Laing, T., & Kang, C. (2015). On normality assumption in residual simulation for probabilistic load forecasting. *IEEE Transactions on Smart Grid*, Advance Online Publication. http://dx.doi.org/10.1109/TSG.2015.2447007.

Xie, J., Liu, B., Lyu, X., Hong, T., & Basterfield, D. (2015). Combining load forecasts from independent experts. In *47th North American power symposium at Charlotte, NC, October 4–6, 2015*.

Jingrui Xie received her bachelor degree in Finance from Sun-Yat Sen University, Guangzhou, China, her masters degree in Economics from Duke University, Durham, NC, USA, and her masters degree in Engineering Management from University of North Carolina at Charlotte, Charlotte, NC, USA, in 2009, 2011 and 2015, respectively. She is currently a Ph.D. student at University of North Carolina at Charlotte and a Senior Associate Research Statistician Developer at SAS Institute Inc. Jingrui was the primary statistician developer for the SAS Energy Forecasting Solution and has provided consulting services on load forecasting to several utilities for system operations, planning and energy trading.

Tao Hong is the Director of BigDEAL (Big Data Energy Analytics Laboratory), NCEMC Faculty Fellow of Energy Analytics, Graduate Program Director, Assistant Professor of Systems Engineering and Engineering Management, and Associate of Energy Production and Infrastructure Center at University of North Carolina at Charlotte. He is the Founding Chair of the IEEE Working Group on Energy Forecasting, General Chair of the Global Energy Forecasting Competition, lead author of the online book *Electric Load Forecasting: Fundamentals and Best Practices*, and author of the blog Energy Forecasting. Dr. Hong received his B.Eng. in Automation from Tsinghua University in Beijing and his Ph.D. with co-majors in Operations Research and Electrical Engineering from North Carolina State University.

Please cite this article in press as: Xie, J., & Hong, T., GEFCom2014 probabilistic electric load forecasting: An integrated solution with forecast combination and residual simulation. International Journal of Forecasting (2015), http://dx.doi.org/10.1016/j.ijforecast.2015.11.005