Global Energy Forecasting Competition 2012

Tao Hong, Pierre Pinson and Shu Fan

Abstract

The Global Energy Forecasting Competition (GEFCom2012) attracted hundreds of participants worldwide, who contributed many novel ideas to the energy forecasting field. This paper introduces both tracks of GEFCom2012, hierarchical load forecasting and wind power forecasting, with details on the aspects of the problem, the data, and a summary of the methods used by selected top entries. We also discuss the lessons learned from this competition from the organizers' perspective. The complete data set, including the solution data, is published along with this paper, in an effort to establish a benchmark data pool for the community.

1. Background

In a broad sense, energy forecasting covers a wide range of forecasting problems in the utility industry, such as generation forecasting, load forecasting, price forecasting, demand response forecasting, and so on. While the deployment of smart grid technologies offers the utility industry data of a higher granularity than ever before, it also presents the challenge of obtaining business value from big datasets. As a result, energy forecasting, one of the most fundamental and classical problems, has found a new life in today's utility industry.

Although a significant amount of the literature has been devoted to energy forecasting, most such studies are still at the theoretical level, having little practical value. No formal benchmarking process or data pool has been established in the field, and new publications rarely reproduce the results from past work done by other research groups for a comparison. Few academic programs in electrical engineering, statistics or economics offer courses which concentrate on energy forecasting. Given these facts, the IEEE Working Group on Energy Forecasting (WGEF) organized the Global Energy Forecasting Competition 2012 (GEFCom2012) in order to (i) improve the forecasting practices of the utility industry, (ii) bring together state-of-the-art techniques for energy forecasting, (iii) bridge the gap between academic research and industry practice, (iv) promote analytics in power and energy education, and (v) prepare the industry to overcome the forecasting challenges posed by the smart grid world. The competition included two tracks, hierarchical load forecasting and wind power forecasting. In this paper, we introduce GEFCom2012 in detail, as well as publishing the complete competition dataset in an attempt to establish a benchmarking data pool for energy forecasting.

We started planning the competition in late 2011; this mainly involved identifying field interest, seeking sponsorships, and setting up the rules and schedule. Most previous forecasting competitions have used a centralized communication approach, where the participants were able communicate with the administrators but not with each other. As a result, the participants did not know the scores and ranks until the administrators calculated them after the competition. The tourism competition (Athanasopoulos, Hyndman, Song, & Wu. 2011) took a different approach, by using Kaggle's platform, where both the participants and administrators can share questions, ideas and findings with each other on Kaggle's forum. As soon as a team submits its entry, the score is calculated and displayed to the team automatically. If the score is the best one presented by this team, the public leader board is refreshed to reflect the changes. Based on these key features, GEFCom2012 selected Kaggle as the competition platform, becoming the second forecasting competition to be hosted by Kaggle. The first Call for Participants was issued in May 2012. Prior to the launching date, we received registrations from around 120 people from over 30 countries. The competition was active on Kaggle for two months, from 8/31/2012 to 10/31/2012, and by the end of the competition, the data on each track had been downloaded by over 600 unique users.

The remainder of this paper is organized as follows: Sections 2 and 3 introduce the two tracks, respectively, in terms of the problem, the data, and a brief summary of the methods and results. Section 4 discusses the issues with and lessons learned from this competition. The paper concludes in Section 5 with an outlook of potential future work. We also acknowledge the key contributors at the end.

2. Hierarchical load forecasting

2.1. Problem description

Short term load forecasting (STLF) provides load forecasts at hourly or sub hourly intervals for the following one day to two weeks. The forecasts are used by all sectors of the utility industry, from generation and transmission to distribution and retail. The reasons why businesses need short term load forecasts include unit commitment, T&D (transmission and distribution) operations and maintenance, and energy market activities. Many different statistical and artificial intelligence techniques have been applied to STLF over the past three decades, such as multiple linear regression (MLR), the Box-Jenkins approach, Artificial Neutral Networks, etc. A comprehensive review of the literature is provided by Hong (2010).

In the hierarchical load forecasting track, the participants were required to backcast and forecast hourly loads (in kW) for a US utility with 20 zones at both the zonal (20 series) and system (sum of the 20 zonal level series) levels, with a total of 21 series. We provided the participants with 4.5 years of hourly load and temperature history data, with eight non-consecutive weeks of load data removed. The backcasting task is to predict the loads of these eight weeks in the history, given actual temperatures, where the participants are permitted to use the entire history to backcast the loads. The forecasting task is to predict the loads for the week immediately after the 4.5 years of history without the actual temperatures or temperature forecasts being given. This is designed to mimic a short term load forecasting job, where the forecaster first builds a model using historical data, and then develops the forecasts for the next few days. Traditionally, most STLF jobs are conducted using system level data only. In this competition, we also provided zonal level data, in order to further mimic a STLF job in the smart grid era, where the forecasters have access to the smart meter information.

Of the thousands of papers in the load forecasting literature, most are devoted to a range of modeling techniques, while many practical issues still have not received enough attention. When designing the competition problem, we wanted to highlight a few challenges, with the aim of encouraging new ideas on the following aspects:

1) Data cleansing. The competition data are real-world data, and include significant data quality issues due to outages, load transfers and various other data errors. An effective data cleansing method

would be expected to enhance the forecasting accuracy. This challenge also applies to the wind forecasting track.

2) Hierarchical forecasting. Different zones have different electricity consumption behaviors. For instance, Zone 9 represents an industrial customer load, which is largely not weather sensitive. In order to utilize the hierarchical information fully, the participants may choose a bottom-up, middle-out or top-down approach. In addition, to avoid the possibility of some participants using additional external data, we did not specify the locations of the zones and weather stations. Therefore, another challenge is to decide which weather station(s) should be associated with each delivery point. In practice, although the forecasters do have access to the geographical information, they still need to decide which weather station(s) should be used for each zone and how to use them.

3) Special days forecasting. The loads of holidays and the surrounding days are usually less predictable than those of regular days, due to the limited sample sizes and the variability of the pattern over time. When selecting the weeks to be backcasted and forecasted, we included holidays in some of the weeks.

4) Temperature forecasting. In an operations environment, some utilities purchase commercial weather forecasts, while others have their own meteorologists and develop in-house weather forecasts. In this competition, we did not release the temperature forecasts for the week to be forecasted. If the participants decided to use temperature variables, they had to develop their own temperature forecast for the week to be forecasted.

5) Ensemble forecasting. The participants were not restricted to any specific techniques or tools for this competition. We would like to see applications of ensemble forecasting methods in both tracks of GEFCom2012.

6) Integration. A load forecasting job covers a few different tasks, including the ones listed above. The integration of these tasks is another important task. For instance, temperature forecasts, which have low errors overall, but high errors during peak load periods, may not result in useful load forecasts. In this case, a good integration strategy should consider the accuracy of the temperature forecasts when applying load forecasting models. From the reports we received, all of them performed the two tasks (temperature forecasting and load forecasting) separately, and then simply fed the temperature forecasts to the load forecasting model in order to generate the load forecasts.

Other than the standard Kaggle rules, we set up the following two rules:

1) The participants are not allowed to use more weather, load or economy data than has been provided.

2) At each hour, the sum of the zonal level loads should be equal to the system level load.

The error score in the hierarchical load forecasting track is the Weighted Root Mean Square Error (WRMSE), given by:

$$WRMSE = \sqrt{\frac{\sum_i w_i (A_i - P_i)^2}{\sum_i w_i}},$$

where A_i and P_i are the actual and predicted values of observation i, while the weight for this observation is denoted as w_i, and specified in Table 1.

Table 1. Weight assignment.

Week(s)	Weight
Forecasted week at system level	160
Forecasted week at zonal level	8
Backcasted week at system level	20
Backcasted week at zonal level	1

2.2. Data description

The complete dataset can be divided roughly into two parts, based on the different purposes of usage: a training set for model identification and parameter estimation, and an evaluation set for calculating scores. Kaggle selects a random 25% of the evaluation data as the validation set, for calculating public scores, and the remaining 75% forms the test set for calculating private scores. The public scores can be seen by all of the participants and competition administrators throughout the competition, while the private scores are published at the end of the competition. The validation and test data were not released to the participants during the competition; now, however, we are publishing the complete dataset along with this paper, including five spreadsheets in Comma-Separated Values (CSV) format for the hierarchical load forecasting track:

1) Load_history. Hourly load history of 20 zones, from the 1st hour of 2004/1/1 to the 6th hour of 2008/6/30, with the following 8 weeks set to be missing for backcasting purposes: 2005/3/6–2005/3/12, 2005/6/20–2005/6/26, 2005/9/10–2005/9/16, 2005/12/25–2005/12/31, 2006/2/13–2006/2/19, 2006/5/25–2006/5/31, 2006/8/2–2006/8/8, and 2006/11/22–2006/11/28.

2) Temperature_history. The hourly temperature history of 11 weather stations, from the 1st hour of 2004/1/1 to the 6th hour of 2008/6/30.

3) Holiday_list. A list of US Federal holidays from 2004/1/1 to 2008/7/7.

4) Load_benchmark. Predicted hourly loads from 2008/7/1 to 2008/7/7. The weight column shows the weights assigned to different weeks and levels.

5) Load_solution. Actual hourly loads from 2008/7/1 to 2008/7/7. The format is similar to "Load_benchmark". The indicator column shows the way in which we split the solution data in order to calculate the scores for public and private leaderboards.

2.3. Summary of methods and results

The benchmark is created based on a MLR model with an intercept and the following effects, as discussed by Hong (2010):

1) main effects: *Trend* (an increasing normal number assigned to each observation in chronological order), *T* (temperature of the current hour), T^2, T^3, *Month* (a class variable, with 12 levels representing the 12 months of a year), *Weekday* (a class variable, with seven levels representing the seven days of a week), and *Hour* (a class variable, with 24 levels representing the 24 hours of a day).

2) cross effects (interactions): *Hour**Weekday*, *T*Month*, T^2**Month*, T^3**Month*, *T*Hour*, T^2**Hour* and T^3**Hour*.

The parameters are estimated using the 4.5 years of history less the 8 backcasted weeks. For each zone, we build 11 models, one per weather station. The weather station with the best fit is then assigned to the corresponding zone. We predict the 8 weeks of loads using the same model with actual temperatures from the selected weather station. We forecast the last week of loads using the same model with forecasted temperatures, where the temperature forecast at each hour is the average temperature at the same date and hour over the past four years.

Table 2 summarizes the methods used by selected entries based on their reports. We also calculate the WRMSEs of the 8 backcasted weeks, 7/1/2008, the entire forecasted week, the validation data, the test data, and all data, as is shown in Table 3, together with the number of submissions each team made.

Table 2. Summary of methods in the hierarchical load forecasting track.

Kaggle ID	**Techniques**	**Data cleansing**	**Weather station selection**	**Holiday effect**	**Temperature forecast**	**Ensemble forecasting**
CountingLab	MLR, singular value decomposition	Yes	11 models corresponding to the 11 weather stations were built	Yes	Using the average temperature of the same hour from similar days in the previous years	Combine forecasts from the 5-best fitted models
James Lloyd	Gradient boosting machines, Gaussian process regression, MLR	Not discussed	Temperatures from all stations were used	No	Estimating the smooth trend and daily periodicity of temperature separately	Combine forecasts from three models
Tololo (EDF)	Semi-parametric regression, with B-splines or cubic regression splines as smooth function	Not discussed	A stepwise procedure was used for each zone to select the station that minimized forecasting error on a test set	Yes	Not discussed	No
TinTin	Nonparametric additive models with P-spline, component-wise gradient boosting	Yes	A testing week (the last week of the available data) was used to determine the station for each zone	Yes	Using the average temperatures at the same period across the previous years	No
Quadrivio	MLR	Yes	Load was fitted to temperature at each	No	Averaging the temperatures during the	No

			station separately, and the best three were used for each zone		same days from previous years	
Chaotic Experiments	Random forest, geometric Brownian motion models	Not discussed	Not discussed	Yes	Not discussed	Combine forecasts from three models
Andrew L	Generalized additive model, spline, PCA	Not discussed	The first component of PCA was used as temperature variable for each hour	No	Using a generalized additive model	No
NHH	Wavelet decomposition, mutual information, neural networks	Not discussed	Temperatures from all stations were considered as input candidate	No	Not discussed	No
TheJellyTeam	Neural networks	Not discussed	Temperatures from all stations were considered	Yes	Using the mean of the same period from the previous years	No
Shooters Touch	Regression models and neural network	No	Weighted average of up to 3 stations, selected based on the fitted result for each station	Yes	Not discussed	No
Tao's Vanilla Benchmark	MLR	No	Best fit from the 11 weather stations	No	Average of the same date/time of the past four years	No

Table 3. Error statistics (WRMSEs) of selected entries in the hierarchical load forecasting track.

Kaggle ID	Backcast	1 day ahead	1 week ahead	Validation	Test	All	Submissions
CountingLab	61890	72504	73900	70700	67215	68160	33
James Lloyd	58406	59273	82346	71164	71467	71387	52
Tololo (EDF)	46756	52136	82776	52669	71780	67223	39
TinTin	50926	112410	86590	64352	73307	71033	42
Quadrivio	71663	63186	81645	72825	78196	76816	29
Chaotic Experiments	78238	50967	89783	93045	80763	84209	19
Andrew L	68638	133005	106272	101069	84850	89456	3
NHH	65360	121818	109850	93641	89174	90385	18
TheJellyTeam	72197	120752	101066	83916	89202	87826	12
Tao's Vanilla Benchmark	69557	148352	123758	112547	95588	100385	1

3. Wind power forecasting

3.1. Problem description

Given the ever-increasing deployment of wind power capacities as a viable renewable energy solution in the electricity mix, a number of decision-making problems in connection with power system operations and a participation in electricity markets require some form of forecasts as input. The development of methods for wind power forecasting can be traced back to the work of Brown, Katz, and Murphy (1984), who used simple time series models for wind forecasting at a site of interest, then converted the resulting wind forecasts to electric power generation by passing them through a theoretical manufacturer's power curve. Since then, three decades of research and development have led to the proposal of a wide range of approaches, with a clear intensification of these efforts since the beginning of the new millennium, as wind power capacities began spreading round the world to a greater extent (previously, they were concentrated mainly in the European region). A set of reviews of the state of the art in wind power forecasting exists, to which the readers are referred for an exhaustive coverage of the alternative approaches. The most complete of these reviews are those by Giebel, Brownsword, Kariniotakis, Denhard, & Draxl, (2011) and Monteiro, Bessa, Miranda, Botterud, Wang, & Conzelmann, (2009).

In the wind power forecasting track, the participants were required to forecast the hourly wind power generation for seven wind farms. We provided three years of historical data, including both wind power generation and wind forecasts. The error score for the wind power forecasting track is the Root Mean Square Error (RMSE). Similar to the hierarchical load forecasting track, in addition to new techniques, we also anticipated some novel ideas in relation to data cleansing, ensemble forecasting and integration.

3.2. Data description

In the wind power forecasting track, we used about three years of data on seven wind farms from the same region of the world as a basis for the design of the competition problem. The data consist of historical power measurements for these wind farms, as well as meteorological forecasts of the wind components at the levels of these wind farms.

The historical power measurements have an hourly temporal resolution, with a high level of availability over that period and for all of the wind farms. They were normalized by the respective nominal capacities of the wind farms, in order to obtain normalized power values between zero and one, thus allowing the original characteristics of the wind farms to be masked. This also enables a scale-free comparison of the forecasting results for the various wind farms.

Meteorological forecasts were gathered for the zonal (u) and meridional (v) components of surface winds at 10 m above ground level. They were extracted from the archive of the European Centre for Medium-range Weather Forecasts (ECMWF). ECMWF issues high-resolution deterministic forecasts twice a day at 00UTC and 12UTC, with a temporal resolution of between 3 hours and 10 days ahead. In order to match the hourly resolution of the power measurements, also required by most forecast applications, the forecasts were interpolated using cubic splines, so as to have an hourly resolution. Only the first 48 hours of each forecast series were collated in the dataset. Note that these meteorological predictions were also given in the form of wind speeds and directions for those who preferred to use them in such a format.

A number of 48-hour periods with missing power observations are defined for validation and testing purposes. The first one is from 1 January 2011 at 01:00 to 3 January 2011 at 00:00. The second one is from 4 January 2011 at 13:00 to 6 January 2011 at 12:00. Note that, in order to be consistent, only the meteorological forecasts that were relevant for the periods with missing power data, which would be available in practice, were given. Each of these two periods then repeats itself every 7 days until the end of the dataset. For instance, the first repetition of the first period is 8 January 2011 at 01:00 to 10 January 2011 at 00:00. The second repetition of the first period is 15 January 2011 at 01:00 to 17 January 2011 at 00:00. In between periods with missing data, power observations are available for updating the models if necessary.

Along with this paper, we publish the complete dataset in the form of 11 spreadsheets (in comma-separated values CSV format) for the wind power forecasting track:

1) WindPower_train. Hourly wind power observations for the seven wind farms from 2009/7/1 to 2010/12/31 (i.e., the training set), without any holes, except potentially as a result of data quality issues.

2) WindPower_eval. Hourly wind power observations for the seven wind farms from 2011/1/1 to 2012/6/28 (i.e., the evaluation set), with holes for the periods for which the forecasts are expected to be produced, as mentioned above.

3) WindForecasts_wf1, …, WindForecasts_wf7. Wind forecasts for the seven wind farms and for the same period as for the measurements. Forecasts are issued every 12 hours, with a forecast horizon of 48 hours and an hourly temporal resolution.

4) WindPower_benchmark. Predicted hourly wind power at the seven wind farms for the holes in the evaluation set.

5) WindPower_solution. Actual wind power measurements for the holes defined in the evaluation set. The format is similar to "WindPower_benchmark". The indicator column shows how we split the solution data when calculating the scores for the public and private leaderboards.

3.3. Summary of methods and results

The persistence method, as one of the simplest approaches to issuing wind power forecasts for these wind farms, is used here as a benchmark. This forecasting approach is based on a random walk model, where the forecasted value is defined as the most recent available observation. The methods used by nine selected teams together are summarized in Tables 4. We also show the error statistics of these nine teams and the persistence benchmark in Table 5, together with the number of submissions made by each team. The error statistics (in RMSEs) are broken down by wind farms, validation data, test data and all data.

Table 4. Summary of methods used in the wind power forecasting track.

Kaggle ID	**Technique**	**Data Cleansing**	**Ensemble Forecasting**
Leustagos	Linear combination of nine models (regression from meteorological forecasts to power, inter-wind farm	No	Yes

	dependencies, autoregressive components, with different model structures)		
DuckTile	Data cleaning, and then local linear regression with wind forecasts, day and time of the year as inputs	Yes	No
MZ	Linear models estimated with regularized least squares with radial basis functions spanning the space of wind forecasts, and autoregressive features	No	No
propeller	Linear regression from wind forecasts to power measurements, then a nonlinear correction with gradient boosting machines (with optimal inputs identified through cross-validation)	Yes	No
Duehee Lee	Plain combination of a large number of neural networks (52) and Gaussian process models (5), mapping all input data to power measurements	No	Yes
MTU EE5260	Linear regression and neural networks for the conversion of meteorological forecasts to power	No	No
SunWind	Plain combination of a power curve model, an autoregressive model, a local linear regression model, and a support vector machine model	No	Yes
ymzsmsd	Sparse Bayesian learning with input measurements and forecasts from all wind farms	No	No
4138 Kalchas	Regularized kernel-based regression for the conversion of meteorological forecasts to power	No	No
Benchmark	Persistence	No	No

Table 5. Error statistics (RMSEs) of selected entries in the wind power forecasting track.

Kaggle ID	**WF1**	**WF2**	**WF3**	**WF4**	**WF5**	**WF6**	**WF7**	**Validation**	**Test**	**All**	**Submissions**
Leustagos	0.145	0.138	0.168	0.144	0.158	0.133	0.140	0.146	0.146	0.146	37
DuckTile	0.143	0.145	0.172	0.145	0.165	0.137	0.146	0.149	0.147	0.148	82
MZ	0.141	0.151	0.174	0.145	0.167	0.141	0.145	0.148	0.149	0.149	19
propeller	0.144	0.153	0.177	0.147	0.175	0.141	0.147	0.148	0.153	0.152	64
Duehee Lee	0.157	0.144	0.176	0.160	0.169	0.154	0.148	0.155	0.155	0.155	10
MTU EE5260 Forecast Team	0.161	0.172	0.193	0.162	0.192	0.156	0.160	0.166	0.169	0.168	20
SunWind	0.174	0.177	0.193	0.176	0.179	0.157	0.162	0.173	0.171	0.172	26

ymzsmsd	0.163	0.186	0.200	0.164	0.192	0.162	0.167	0.173	0.174	0.174	24
4138 Kalchas	0.180	0.179	0.197	0.175	0.200	0.160	0.165	0.179	0.176	0.177	3
Benchmark	0.302	0.338	0.373	0.364	0.388	0.341	0.361	0.361	0.353	0.355	1

4. Discussion

Figure 1 shows the cumulative number of unique IDs that downloaded the data from each track from the beginning of the competition. The vertical dash-dot line indicates the end of the competition, at which point there were about 600 unique IDs from each track. After the competition, the data were still being downloaded by the Kaggle users. Using Kaggle's platform, the competition attracted many more participants than expected, many of whom were very experienced data scientists outside the utility industry. While the diverse range of backgrounds of the participants introduces a lot of new ideas into the energy forecasting field, some of the participants are not interested in joining in with the post-competition activities, such as submitting reports, presenting their work at conferences, and writing scientific papers.

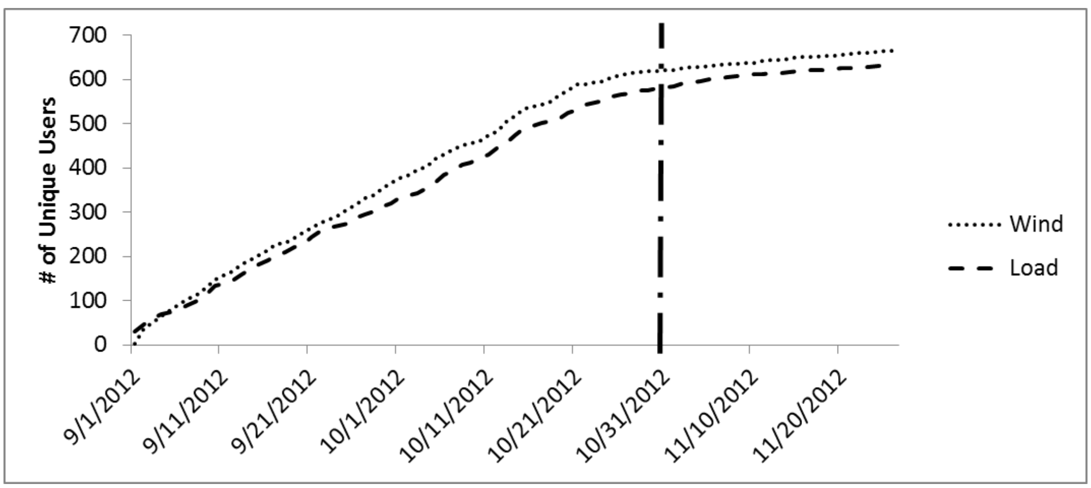

Figure 1. Number of unique users who downloaded competition data from the two tracks during the first 3 months.

Kaggle provides a forum where participants and competition administrators can post questions, answers and findings. This feature allows the participants to help each other in the public domain. It also allows the administrators to address issues as soon as they are raised. As the competition proceeds, there is rich body of information in the forum, which requires the new participants to review the old posts. Some participants, chiefly new Kaggle users, may not review the previous posts, which can lead to violations of some competition rules. In order to avoid similar situations in the future, we would recommend that competition administrators increase the participants' awareness of important posts in the forum discussion.

In the hierarchical load forecasting track, in order to maintain the load level of each zone, we gave the actual loads instead of standardized values, which opens the possibility that some participants may be able to guess the location of the utility, and use external information to win the competition. To avoid this situation, we required the teams to submit reports and codes, which were then evaluated by the award committee of GEFCom2012. Ultimately, two teams were disqualified due to their use of actual temperature data in the forecasted week. In the wind power forecasting track, the data were standardized, so that the participants could not find the solution by guessing where the wind farms are.

In real world short term load or wind forecasting jobs, forecasters have to develop their forecasts on daily basis using newly available information. In other words, the forecast origin moves every day. In order to implement this feature in a competition, we would have to host multiple phases, with new data being released at each phase. While each phase might take a couple of weeks to complete, the entire competition would take much longer than two months. Implementing this feature would also require the participants to be fully engaged throughout the competition. This is more achievable as an in-class competition than an inaugural international competition, and therefore, we did not set up this feature when designing GEFCom2012. As an amendment, we leave a few missing periods in the history for prediction. Since we cannot really determine whether the participants are using data after a missing period when predicting this missing period, we did not restrict the participants to using only the data prior to each missing period being predicted. This setup may mean that regression or some other data mining techniques have an advantage over some time series forecasting techniques such as ARIMA, which may be part of the reason why we did not receive any reports using the Box-Jenkins approach in the hierarchical load forecasting track.

By nature, forecasting is a stochastic problem. In the utility industry, some applications in some utilities require probabilistic forecasts in the form of predictive densities or scenarios as inputs, such as annual peak demand forecasting for system planning (Hyndman & Fan, 2012), systems reserve quantification (Matos & Bessa, 2011), unit commitment (Tuohy, Meibom, Denny, & O'Malley, 2009), and trading of wind power generation (Pinson, Chevallier, & Kariniotakis, 2007). On the other hand, a lot of decision-making processes are set to take point forecasts only. The majority of the energy forecasting literature has considered point forecasts. In GEFCom2012, in order to keep the competition problem and error scores straightforward, we let the participants develop point forecasts rather than probabilistic ones.

5. Conclusion

GEFCom2012 includes two tracks: hierarchical load forecasting and wind power forecasting. The competition attracted hundreds of participants worldwide. In this paper, we have introduced GEFCom2012 from several aspects, including the background, problem, data, methods, results, and lessons learned. We have also published the complete dataset from each track in an attempt to establish a data pool for energy forecasting. In the future, we would like to expand the competition by adding more tracks, such as long term load forecasting, price forecasting and solar generation forecasting. We would also like to explore other features, such as a rolling forecast origin, comprehensive error scores, and probabilistic forecasts.

Acknowledgement

The authors gratefully acknowledge the following organizations, which contributed to the success of GEFCom2012: IEEE Power& Energy Society, IEEE Power Systems Planning and System Implementation Committee, IEEE Power and Energy Education Committee, IEEE Working Group on Energy Forecasting, Kaggle, WeatherBank Inc., *International Journal of Forecasting*, and IEEE Transactions on Smart Grid. The authors also appreciate support in organizing GEFCom2012 from the following individuals: David Hamilton, Eric Wang, Hamidreza Zareipour, ML Chan, Rob J Hyndman, Wei-Jen Lee, Fangxing Li, Shanshan Liu, Anil Pahwa, Mohammad Shahidehpour, and Kumar Venayagamoorthy.

References

Athanasopoulos, G., Hyndman, R., Song, H., & Wu, D. (2011). The tourism forecasting competition. *International Journal of Forecasting*, *27*(3), 822–844.

Brown, B. G., Katz, R. W., & Murphy, A. H. (1984). Time series models to simulate and forecast wind speed and wind power. *Journal of Climate and Applied Meteorology*, *23*, 1184-1195.

Hyndman, R.J., & Fan, S. (2009) Density forecasting for long-term peak electricity demand. ***IEEE Transactions on Power Systems***, **25**(2), 1142-1153.

Giebel, G., Brownsword, R., Kariniotakis, G., Denhard, M., & Draxl, C. (2011). *The state-of-the-art in short-term prediction of wind power: A literature overview* (2nd ed.). Tech. Rep. EU project ANEMOS.plus, available at: http://orbit.dtu.dk.

Hong, T. (2010). *Short term electric load forecasting*. North Carolina State University.

Matos, M. A., & Bessa, R. J. (2011). Setting the operating reserve using probabilistic wind power forecasts. *IEEE Transactions on Power Systems*, *26*, 594-603.

Monteiro, C., Bessa, R., Miranda, V., Botterud, A., Wang, J., & Conzelmann, G. (2009). *Wind power forecasting: State of the art 2009*. Tech. Rep., Argonne National Laboratory, Decision and Information Sciences Division, ANL/DIS-10-1.

Pinson, P., Chevallier, C., & Kariniotakis, G.N. (2007). Trading wind generation from short-term probabilistic forecasts of wind power. *IEEE Transactions on Power Systems*, *22*, 1148-1156.

Tuohy, A., Meibom, P., Denny, E., & O'Malley, M. (2009). Unit commitment for systems with significant wind penetration. *IEEE Transactions on Power Systems*, *24*, 592-601.

Bio

Tao Hong received his B.Eng. in Automation from Tsinghua University, Beijing, and an M.S. in Electrical Engineering, an M.S. in Operations Research and Industrial Engineering, and a Ph.D. in Operations Research and Electrical Engineering from North Carolina State University. He is the Head of

Energy Forecasting at SAS Institute Inc., Chair of IEEE Working Group on Energy Forecasting, General Chair of the Global Energy Forecasting Competition, and Guest Editor-in-Chief of the *IEEE Transactions on Smart Grid*'s Special Issue on Analytics for Energy Forecasting with Applications to Smart Grid. His research interests are in analytics and utility applications.

Pierre Pinson holds a M.Sc. in Applied Mathematics, as well as a Ph.D. in Energy from Ecole des Mines de Paris, France. He is the Associate Professor in Stochastic Energy Systems at the Technical University of Denmark, with particular interests in probabilistic forecasting for energy and electricity markets. He is involved as principal scientist and workpackage leader in a number of Danish and European projects related to these topics. He acts as editor for the journals *Wind Energy* and *IEEE Transactions on Power Systems*.

Shu Fan received his B.S., M.S., and Ph.D. degrees in the Department of Electrical Engineering from China's Huazhong University of Science and Technology (HUST), Wuhan, China, in 1995, 2000, and 2004, respectively. He conducted postdoctoral research sponsored by the Japanese Government in Osaka Sangyo University from 2004 to 2006, and was a Visiting Assistant Professor at the Energy Systems Research Center at the University of Texas at Arlington from 2006 to 2007. At present, he is a Senior Research Fellow at Monash University, Clayton, Australia. His research interests include energy system forecasting, power system control, and high-power power electronics.