ARTICLE IN PRESS
International Journal of Forecasting ▮ (▮▮▮▮) ▮▮▮–▮▮▮
Contents lists available at ScienceDirect
International Journal of Forecasting
journal homepage: www.elsevier.com/locate/ijforecast

Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond

Tao Hong[a,*], Pierre Pinson[b], Shu Fan[c], Hamidreza Zareipour[d], Alberto Troccoli[e], Rob J. Hyndman[c]

[a] University of North Carolina at Charlotte, USA
[b] Technical University of Denmark, Lyngby, Denmark
[c] Monash Business School, Monash University, Clayton, Australia
[d] University of Calgary, Alberta, Canada
[e] World Energy & Meteorology Council, UK

ARTICLE INFO

Keywords:
Electric load forecasting
Electricity price forecasting
Wind power forecasting
Solar power forecasting
Probabilistic forecasting
Forecasting competition

ABSTRACT

The energy industry has been going through a significant modernization process over the last decade. Its infrastructure is being upgraded rapidly. The supply, demand and prices are becoming more volatile and less predictable than ever before. Even its business model is being challenged fundamentally. In this competitive and dynamic environment, many decision-making processes rely on probabilistic forecasts to quantify the uncertain future. Although most of the papers in the energy forecasting literature focus on point or single-valued forecasts, the research interest in probabilistic energy forecasting research has taken off rapidly in recent years. In this paper, we summarize the recent research progress on probabilistic energy forecasting. A major portion of the paper is devoted to introducing the Global Energy Forecasting Competition 2014 (GEFCom2014), a probabilistic energy forecasting competition with four tracks on load, price, wind and solar forecasting, which attracted 581 participants from 61 countries. We conclude the paper with 12 predictions for the next decade of energy forecasting.

1. Introduction

In today's competitive and dynamic environment, the energy supply, demand and prices are becoming increasingly volatile and unpredictable. More and more decision-making processes in the energy industry require a comprehensive outlook of the uncertain future. Many decision makers are relying on probabilistic forecasts to quantify these uncertainties, rather than point forecasts.

Here, we use the term "energy forecasting" to refer to "forecasting in the energy industry", which includes but is not limited to the forecasting of the supply, demand and price of electricity, gas, water, and renewable energy resources. Probabilistic forecasts can take various forms, e.g., from quantile to full density forecasts, and probabilistic forecasts for multi-categorical variables or functional data. The business needs for probabilistic energy forecasts spread across the planning and operations of the entire energy value chain.

Thousands of papers on energy forecasting have been published over the past half-century. Hong (2014) provided an overview of energy forecasting, tracing the forecasting practices back to the inception of the electric power industry. A few recent literature review articles have offered more comprehensive views for various subdomains of energy forecasting, such as wind power forecasting (Pinson, 2013; Zhang, Wang, & Wang, 2014), electric load forecasting (Hong & Fan, in this issue), and electricity price forecasting (Weron, 2014). While there

* Corresponding editor.
E-mail address: hong@uncc.edu (T. Hong).

http://dx.doi.org/10.1016/j.ijforecast.2016.02.001
0169-2070/

are differences among the subdomains, there are some common challenges:

(1) Data cleansing. The real-world data for energy forecasting are not always clean. Data cleansing was recognized as one of the challenges of the Global Energy Forecasting Competition 2012 (GEFCom2012), which focused on point forecasting (Hong, Pinson, & Fan, 2014). Data cleansing remains a challenge for probabilistic energy forecasting.
(2) Probabilistic forecasting methodologies. Different subdomains in energy forecasting have different levels of maturity in their probabilistic forecasting. Which probabilistic forecasting methods, such as the ones reviewed by Gneiting and Katzfuss (2014), are applicable to energy forecasting? What specific methodologies are most suitable for a given subdomain? What is the best way to generate input scenarios? Which techniques are most effective at generating probabilistic forecasts? How to simulate residuals?
(3) Forecast combination. Combining a set of point forecasts usually results in more accurate and robust point forecasts. Can we adopt a similar concept for probabilistic energy forecasting? Should we combine point forecasts to give one new point forecast first, then generate a probabilistic forecast, or should we combine point forecasts to give a probabilistic forecast directly? How should probabilistic forecasts be combined to generate a better probabilistic forecast?
(4) Integration. A probabilistic forecasting process can be dissected into several components, such as scenario generation, modeling, and post-processing. An optimal outcome from one component may not be the optimal one for the entire process. The challenge is to integrate the various steps in order to obtain a high quality probabilistic forecast. A similar challenge of integration was also recognized in GEFCom2012 (Hong, Pinson, & Fan, 2014).

In order to maintain the momentum initiated by GEFCom2012, stimulate research activity, and tackle challenges in probabilistic energy forecasting, we decided to start two initiatives simultaneously: (1) organizing a special section for the *International Journal of Forecasting* on probabilistic energy forecasting; and (2) organizing GEFCom2014 with the plan to include the winning methodologies in the same special section.

A call for papers on probabilistic energy forecasting was released on October, 2013. We received 34 submissions, of which seven were accepted for publication in this special section. In addition, we also collected 13 papers from the top entries of GEFCom2014 and one paper from the winning entry of an in-class probabilistic load forecasting competition following a setup similar to that of the probabilistic load forecasting track of GEFCom2014. In total, this special section has collected 21 papers in addition to this hybrid editorial and review paper.

This paper serves four purposes: (1) to discuss the seven non-GEFCom2014 papers collected for this special issue; (2) to introduce the GEFCom2014 and the winning methodologies; (3) to provide an outlook for the field of probabilistic energy forecasting; and (4) to publish the data related to GEFCom2014 and the in-class competition. Section 2 summarizes the non-GEFCom2014 papers. Section 3 discusses the organization of GEFCom2014. Sections 4–7 introduce the four tracks of GEFCom2014, including the problem, the data, and the methods followed by the winning teams. The paper concludes in Section 8 with an outlook for the next decade of probabilistic energy forecasting.

2. Non-GEFCom2014 papers

The seven non-GEFCom2014 papers include three on demand forecasting, two on price forecasting and two on renewable generation forecasting. The subjects being forecasted include electricity and gas demand, electricity prices, wind speed and wave energy. Table 1 lists the authors and titles of the papers.

2.1. Electricity and gas demand forecasting

The majority of the load forecasting literature has focused on point forecasting. Probabilistic load forecasting (PLF) has become attractive to the load forecasting community only over the last decade. In this issue, Hong and Fan offer a tutorial review on PLF. Because most of the studies in the PLF literature have been developed from point load forecasting techniques and methodologies, this tutorial review begins by covering a selection of papers on point load forecasting. The authors then review the research progress on PLF made by two groups, the business consumers of load forecasts and the load forecasters. After reviewing the point and probabilistic load forecasting literature, the authors dissect the PLF problem into three elements, namely the input, model and output. They then introduce ways of producing probabilistic load forecasts from each element and evaluating the probabilistic load forecasts. Finally, the authors conclude the review with an in-depth discussion of future research needs. In addition to their review of the probabilistic load forecasting literature and tutorial about the production and evaluation of probabilistic load forecasts, the authors also offer their opinions about the myth of best techniques, the novelty and significance of load forecasting research, and the importance of integration in load forecasting.

Antoniadis et al. propose a flexible nonparametric function-valued forecasting model. The predictor can be viewed as a weighted average of the futures of past situations, where the weights increase with the similarity between the past situations and the actual one. This strategy can provide simultaneous point predictions at multiple horizons. In addition, these weights also induce a probability distribution that can be used to produce bootstrap pseudo-predictions. Prediction intervals can then be constructed after obtaining the corresponding bootstrap pseudo-prediction residuals. In their paper, Antoniadis et al. propose to obtain prediction intervals that are valid simultaneously for the multiple prediction horizons that correspond to the relevant path forecasts. The methodology is demonstrated using a dataset from the French grid.

Anomaly detection is an important step in data analysis, and has been regarded as one of the challenges of both

Please cite this article in press as: Hong, T., et al., Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. International Journal of Forecasting (2016), http://dx.doi.org/10.1016/j.ijforecast.2016.02.001

Table 1
Authors and titles of the seven non-GEFCom2014 papers collected in this special section.

Authors	Paper title
T. Hong and S. Fan	Probabilistic electric load forecasting: a tutorial review
A. Antoniadis, X. Brossat, J. Cugliari, J. Poggi	A prediction interval for a function-valued forecast model: application to load forecasting
H. Akouemo, R. Povinelli	Probabilistic anomaly detection in natural gas time series data
K. Maciejowska, J. Nowotarski, R. Weron	Probabilistic forecasting of electricity spot prices using factor quantile regression averaging
A. Bello, J. Reneses, A. Muñoz, A. Delgadillo	Probabilistic forecasting of hourly electricity prices in the medium-term using spatial interpolation techniques
E. Iversen, J. Morales, J. Møller, H. Madsen	Short-term probabilistic forecasting of wind speed using stochastic differential equations
J. Jeon, J. Taylor	Short-term density forecasting of wave energy using ARMA-GARCH models and kernel density estimation

GEFCom2012 (Hong, Pinson et al., 2014) and GEFCom2014. There are many types of anomalies in the gas load forecasting field, and not all of them can be detected using common practices such as summary statistics. Akouemo and Povinelli examine and categorize a wide range of factors that can induce anomalies. They then propose an anomaly detection method based on linear regression models and a Bayesian maximum likelihood classifier for natural gas time series. The gas load data first flows through a regression model that helps to identify anomaly candidates. The candidates are then tested for false positives and classified using a Bayesian classifier. The case study is performed on a daily gas consumption series. Note that a winning team in GEFCom2014, Jingrui Xie, followed a similar regression-based strategy for data cleansing.

2.2. Electricity price forecasting

Although the forecasting community is well aware that taking the simple average of different forecasts often results in forecasts that are more accurate and robust, forecast combination is not seen often in the electricity price forecasting literature. Maciejowska et al. take forecast combination to another level by averaging day-ahead point forecasts in order to derive probabilistic price forecasts. The core methodology involves using quantile regression to average large numbers of point forecasts, using principal component analysis (PCA) to extract the major factors driving the individual forecasts and hence automate the process of selecting forecasts to be included in the combination. The case study is performed on data from the British power market. Based on a comprehensive evaluation using the unconditional coverage, the conditional coverage and the Winkler score, the proposed method outperforms both the benchmark autoregressive exogenous (ARX) model and the quantile regression averaging (QRA) without PCA.

Most electricity price forecasting studies in the literature have focused on short horizons. Very little research has been conducted into medium-term price forecasting. Research on medium-term probabilistic price forecasting is even more rare. Bello et al. propose a methodology that combines a market equilibrium model with a Monte Carlo simulation, integrated with spatial interpolation techniques and a new definition of load levels, for medium-term hourly probabilistic price forecasting. The authors demonstrate the effectiveness of the proposed methodology using public data from the Spanish market. As is mentioned in the paper, the proposed methodology is currently used by a major Spanish electricity company.

2.3. Renewable generation forecasting

Most of the papers on probabilistic wind power forecasting literature over the last five years or so have focused on different variants of statistical and machine learning approaches, generalized to generate probabilistic forecasts. Very little attention has been paid to the potential of employing stochastic differential equations for that purpose, even though there may be obvious advantages of placing ourselves in a continuous formulation framework. This is because, for the same process, the forecasting models based on stochastic differential equations are necessarily simpler and more compact than the counterpart. Similarly, the same model can be used for generating forecasts at various resolutions, through numerical integration methods. Finally, the same modeling approach may be used for issuing forecasts in various forms, from traditional single-valued predictions to trajectories that describe both the characteristics of marginals at each lead time and the temporal dependence structure in the power generation dynamics. In the present case, Iversen et al. propose the use of stochastic differential equations for describing the conversion of the information given by weather forecasts into the power production observed at a wind farm. An interesting method proposed by the authors consists of considering the dynamics (e.g., the gradient) in the input information, hence capturing the wind power dynamics better. Disregarding such dynamics would yield forecasts that would describe the variability in wind power generation fluctuations only poorly.

The paper by Jeon and Taylor concentrates on an original wave energy forecasting problem. This type of problem has seldom been dealt with in the scientific literature, probably because wave energy has not been developed as fast or as strongly as wind or solar power. In a way, wave energy forecasting could be seen as relatively easier, since one expects more regularity in waves (and therefore smoothness and predictability in the wave and wave energy time series). However, waves are the result of both large scale (e.g., swell) and local (e.g., surface winds) effects, which may actually result in complex features in wave time series. In addition, the conversion of wave characteristics into wave energy is sophisticated, since it involves wave heights and periods, and water densities. Disregarding this last variable, which should be easier to handle, in order to deduce the wave energy flux one should still be predicting the wave period and height jointly. This motivates the proposal and evaluation of the bivariate

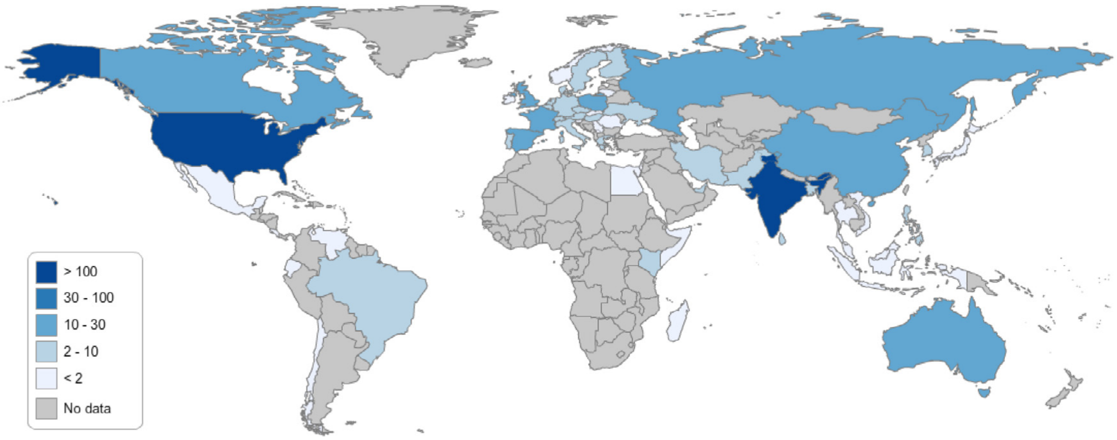

Fig. 1. Geographic distribution of GEFCom2014 contestants (581 people from 61 countries).

modelling and forecasting approaches considered in that paper. The authors also apply it in another context, joint wind and wave forecasting, which may be of great interest in the future if one were to consider having offshore energy hubs where energy would be collected from both wind and waves. From a methodological point of view, the authors build on some of their previous work, using ARMA-GARCH models for the dynamics of the underlying physical processes, combined with conditional kernel density estimation for obtaining probabilistic forecasts.

3. Global Energy Forecasting Competition 2014

3.1. Planning, important dates and participation

The IEEE Power and Energy Society approved financial support for GEFCom2014 at the end of October 2013. At the same time, we (the six authors of this paper: Tao Hong, Pierre Pinson, Shu Fan, Hamidreza Zareipour, Alberto Troccoli, and Rob J. Hyndman) formed the executive committee of GEFCom2014, with Tao Hong as the General Chair. The formal planning of GEFCom2014 started immediately.

Looking back at GEFCom2012, we acknowledged that one of the success factors was the interactive competition platform, where contestants could communicate with each other and with the competition organizers. We decided to maintain this key feature when building GEFCom2014. Meanwhile, we upgraded the competition with three new features:

(1) The probabilistic forecasting theme, to better capture the uncertainties in the modern grid;
(2) Four tracks, on forecasting the electric load (GEFCom2014-L), electricity price (GEFCom2014-P), and wind (GEFCom2014-W) and solar (GEFCom2014-S) power, respectively; and
(3) Rolling forecasting, with incremental data released on a weekly basis for 15 weeks, to mimic real world forecasting processes.

However, the platform used in GEFCom2012 did not offer the feature of rolling forecasting off-the-shelf, nor did it offer proper scoring rules for evaluating probabilistic forecasts. Thus, to accommodate the new features in GEFCom2014, we worked with CrowdANALYTIX to develop a state-of-the-art forecasting competition platform.

One lesson learned from GEFCom2012 was that many active users of Kaggle have expertise in data mining, but not extensive experience with forecasting or energy. In order to attract contestants with more relevant background, we decided to take advantage of the social networks of the competition organizers and contestants. Instead of using the community provided by CrowdANALYTIX, we chose LinkedIn as the communication platform. Prior to the launch date of August 15th, 2015, we received registrations from over 200 people across over 40 countries. The competition was active on CrowdANALYTIX for four months (8/15/2014–12/15/2014). By the end of the competition, the four tracks (load, price, wind and solar) had attracted 362, 287, 254 and 250 contestants, respectively. Some contestants joined multiple tracks. In total, 581 contestants from 61 countries joined GEFCom2014. Fig. 1 shows the distribution of contestants across the globe.

To help the contestants grasp the representative methods for probabilistic energy forecasting quickly, we recommended 11 research papers. Since probabilistic energy forecasting can be seen as the intersection between probabilistic forecasting and energy forecasting, we recommended two reviews, one on probabilistic forecasting (Gneiting & Katzfuss, 2014) and the other on energy forecasting (Hong, 2014). We also recommended the paper that introduced the GEFCom2012 (Hong, Pinson et al., 2014), and one to three papers for each of the tracks, including three for load forecasting (Hong & Fan, in this issue; Hong, Wilson, & Xie, 2014; Hyndman & Fan, 2010); one for price forecasting (Weron, 2014); two for wind forecasting (Pinson, 2013; Zhang et al., 2014); and two for solar forecasting (Bacher, Madsen, & Nielsen, 2009; Lorenz, Scheidsteger, Hurka, Heinemann, & Kurz, 2011).

One of the objectives of GEFCom2014 was to promote education in energy analytics. To encourage students and faculty participation, we decided to offer institute prizes to up to three best-performing academic institutes.

Table 2 lists the important dates of GEFCom2014. The entire competition lasted for 16 weeks. The first month (four weeks, from August 15th to September 13th, 2014)

Please cite this article in press as: Hong, T., et al., Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. International Journal of Forecasting (2016), http://dx.doi.org/10.1016/j.ijforecast.2016.02.001

Table 2
Important dates of GEFCom2014.

Activities	Date
Financial sponsorship confirmed (planning started)	October 31st, 2013
Platform sponsorship confirmed	December 13th, 2013
LinkedIn group started	May 18th, 2014
Historical data released (competition started)	August 15th, 2014
Evaluation period started	September 14th, 2014
Registration deadline	October 10th, 2014
Evaluation period ended	December 6th, 2014
Final report and code due (competition ended)	December 15th, 2014
Finalist presentations at PES General Meeting	July 29th, 2015
GEFCom2014 report submitted to PES	October 6th, 2015

was the trial period. There were three data releases (one release of historical data and two releases of incremental data) during the trial period, and the scores of the trial period were not counted towards the final score. The evaluation period included the 12 weeks from September 14th to December 6th, 2014. At the beginning of each week, we released some incremental data. The scores of the evaluation period were counted towards the final score.

3.2. Error measure

Although the maturity of forecast evaluation differed among the four GEFCom2014 tracks, we decided to settle on a single measure to be used for all four tracks, to keep things simple. We discussed several possible error measures for GEFCom2014, such as the mean absolute error (MAE), the Kolmogorov–Smirnov statistic (K–S statistic), the pinball loss function, and the continuous rank probability score (CRPS). Since the theme was probabilistic energy forecasting, the error measure needed to be a proper scoring rule for probabilistic forecasting. This left the pinball loss function and CRPS as the two most appropriate options. Considering ease of implementation and communication, we decided to use the pinball loss function, which is closely related to the CRPS anyway.

For each time period over the forecast horizon, the participants needed to provide the 1st, 2nd, ..., 99th percentiles, calling these $q_1, \dots, q_{99}$, with $q_0 = -\infty$, or the natural lower bound, and $q_{100} = \infty$, or the natural upper bound. The full predictive densities composed by these quantile forecasts were to be evaluated by the quantile score calculated through the pinball loss function.

For a quantile forecast q_a, with $a/100$ as the target quantile, this score L is defined as:

$$L(q_a, y) = \begin{cases} \left(1 - \dfrac{a}{100}\right)(q_a - y), & \text{if } y < q_a \\ \dfrac{a}{100}(y - q_a), & \text{if } y \geq q_a, \end{cases}$$

where y is the observation used for forecast evaluation, and $a = 1, 2, \dots, 99$.

To evaluate the full predictive densities, this score is then averaged over all target quantiles for all time periods over the forecast horizon and for all zones. A lower score indicates a better forecast.

3.3. Rating and ranking method

When designing this competition, we used a simple method (the trimmed mean) to calculate the final score for each team. This was easy both for CrowdANALYTIX to implement and for the participants to understand. However, applying the trimmed mean method to quantile scores was not comprehensive enough for evaluating rolling forecasts. A major drawback of this simple method was that it treated the point spread as having the same weight in all tasks, without considering that some tasks had more variable scores than others. To fix this, we created a rating that referred to a benchmark score, where the benchmarks were generated using naïve models. In addition, we also wanted to use a scoring method that would give preference to teams that outperformed the benchmark more times, made fewer mistakes, and had stronger performances in the more recent tasks. These considerations led to the rating and ranking methodology below:

Step 0: Initialization. We start with the scores collected over the course of the competition, highlighting the missing entries in blue, and the erroneous entries in yellow. Please refer to the online supplements in Appendix A for the rating and ranking spreadsheets.

Step 1: Ranking permutation. We create a ranking matrix for each valid entry in its corresponding week. For each team, we calculate the second-worst ranking among the valid entries. We permute the rankings of the blue and yellow entries of a team using its second-worst ranking. In addition, the following rules apply:

(1) For a given week, if its second-worst ranking is greater than the number of valid entries, its permuted ranking will be equal to the number of valid entries.
(2) If all of the valid entries corresponding to this team's second-worst ranking are above the benchmark, its permuted ranking will be set to be equal to or better than the ranking of the benchmark.
(3) If some of the valid entries corresponding to its second-worst ranking are worse than the benchmark, we treat missing (blue) and erroneous (yellow) entries differently. For a blue entry, its ranking will be equal to the second-worst ranking. In other words,

Table 3
Winning teams of GEFCom2014.

Track	Ranking	Team name	Team member(s)	Country	Score
Load	1	Tololo	P. Gaillard, Y. Goude, and R. Nedellec	France	50.8%
	2	Adada	V. Dornonnat, A. Pichavant, and A. Pierrot	France	49.8%
	3	Jingrui (Rain) Xie	J. Xie	USA	48.9%
	4	OxMath	G. Giasemidis and S. Haben	UK	48.5%
	5	E.S. Mangalova	E. Mangalova	Russia	46.2%
Price	1	Tololo	P. Gaillard, Y. Goude, and R. Nedellec	France	71.7%
	2	Team Poland	K. Maciejowska and J. Nowotarski	Poland	67.7%
	3	GMD	G. Dudek	Poland	67.1%
	4	C3 Green Team	Z. Kolter, R. Juban, H. Ohlsson and M. Maasoumy	USA	65.0%
	5	pat1	F. Lemke	Germany	64.5%
Wind	1	kPower	M. Landry, T. P. Erlinger, D. Patschke and C. Varrichio	USA	56.6%
	2	dmlab	G. Nagy, G. Borbely, G. Simon and G. Barta	Hungary	55.9%
	3	E.S. Mangalova	E. Mangalova	Russia	55.7%
	4	C3 Green Team	Z. Kolter, R. Juban, H. Ohlsson and M. Maasoumy	USA	55.5%
	5	Yao Zhang	Y. Zhang	China	54.7%
Solar	1	Gang-gang	J. Huang and M. Perry	Australia	68.1%
	2	dmlab	G. Nagy, G. Borbely, G. Simon and G. Barta	Hungary	67.6%
	3	C3 Green Team	Z. Kolter, R. Juban, H. Ohlsson and M. Maasoumy	USA	66.1%
	4	UT_Argonne	D. Lee, Z. Zhou, Y. Kwon	USA	64.3%
	5	Giuseppe C.	G. Casalicchio	Germany	65.4%

its permuted ranking may be below the benchmark. For a yellow entry, its permuted ranking will be equal to or better than the ranking of the benchmark.

Step 2: Score permutation. With the permuted rankings, we can then permute the scores. A permuted score for team i in week j with permuted ranking r is the same as the score from a team with ranking r in week j.

Step 3: Rating and ranking. We define the rating for each entry as "the percentage by which it beats the benchmark". To give preference to teams that improve their methodologies along the way, we assign a linearly increasing weight to the 12 evaluation weeks. The last week is weighted as 12 times the first week. We also make the sum of the 12 weights equal to one. Such a weighting method will also reduce the impact of missing and erroneous entries during the first few weeks. A team's rating is the weighted sum of its ratings over the 12 weeks. This rating tells us roughly how much a team improves on the benchmark. The rankings are calculated based on the ratings in ascending order.

The implementation of the above rating and ranking method is published along with this paper in its online supplements (see Appendix A).

3.4. Winning teams and universities

The leaderboard was updated every week based on the forecasts submitted for each task. The final score from each team was then calculated after the four months of the competition. The participants submitted their reports through CrowdANALYTIX, after which Tao Hong, the General Chair of GEFCom2014, distributed the reports to the other members of the executive committee for review. Having reviewed the reports submitted by the contestants, the executive committee then adjusted the rankings as necessary. This adjustment favored teams that used more practical methods and submitted more comprehensive reports. Table 3 lists the top five teams from each track. Note that UT_Argonne has a lower score than Giuseppe C., but a higher ranking, due to an adjustment based on the quality of their report. This is the only adjustment that was made by the executive committee.

The institute prize winners were determined based on the performances of the participating teams. Each individual team that beat the benchmark and ranked among the top eight of a track received $10 - r$ points, where r is the ranking of the team. A team who participated and ranked in the top eight in multiple tracks received a number of points equal to the sum of the points from each track. The institute score is the sum of the points of all teams associated with the institute. To win, an institute has to have at least two teams beating the benchmark(s) and at least one team in the top eight for a track. Table 4 lists the winning universities and their contributing teams.

4. Probabilistic electric load forecasting

4.1. Problem and data description

The aim of the GEFCom2014-L was to forecast the quantiles of hourly loads for a US utility on a rolling basis. The forecast horizon was one month. Hourly historical load and weather data for the utility were provided. In addition to the data provided by the competition organizer, the contestants were also allowed to use US federal holiday information, including the dates listed in Table 5.

The first data release on August 15th included 69 months of hourly load data (from January 2005 to September 2010) and 117 months of hourly weather data (from January 2001 to September 2010). Starting from the second data release, we made one month of hourly load and weather data available to the contestants each week

Table 4
Winning universities of GEFCom2014 and the contributing teams.

University	Faculty advisor	Institute score	Team name	Track	Ranking
Siberian State Aerospace University, Russia	Olesya Shesterneva	18	E.S. Mangalova	Wind	3
			E.S. Mangalova	Load	5
			Arkadiy Strelnikov	Price	6
			E.S. Mangalova	Price	8
			SAOR	Load	15
			Power Team (SAOR)	Solar	15
University of North Carolina at Charlotte, USA	Tao Hong	14	Jingrui (Rain) Xie	Load	3
			Yanghai Cong	Price	7
			Bidong Liu	Load	8
			Jiali Liu	Wind	8
			Florencio Gonzalez	Price	9
			Ying Chen	Solar	9
			Christopher Benfield	Load	11
			Mohamed Abuella	Solar	12
			Nikolina	Load	13
Tsinghua University, China	Chongqing Kang	6	T_morning	Solar	6
			THU_EILAB#6	Solar	8
			Sniper	Load	10

Table 5
US federal holidays.

Holiday name	Date
New year's day	January 1
Birthday of Martin Luther King, Jr.	Third Monday in January
Washington's birthday (Presidents' day)	Third Monday in February
Memorial day	Last Monday in May
Independence day	July 4
Labor day	First Monday in September
Columbus day	Second Monday in October
Veterans day	November 11
Thanksgiving day	Fourth Thursday in November
Christmas day	December 25

as the solution of the previous week. The load forecasting track involved a total of 11 years of weather data and five years of load data.

Figs. 2 and 3 show the load and temperature series, respectively. Fig. 4 shows a scatter plot of the load and temperature in 2010. The three figures depict the salient features that are typically studied in the load forecasting literature.

In addition to the common challenges mentioned in the introduction section, we also designed the load forecasting track to address a few more:

(1) Weather station selection. We provided 25 weather stations but no identification of their geographical locations. We expected that the contestants would develop some advanced algorithms for selecting weather stations. This was similar to the setup of the hierarchical load forecasting track in GEFCom2012.
(2) Multi-horizon load forecasting. We chose one-month-ahead load forecasting as the competition topic, so that there was some room for the contestants to develop short-term load forecasting models to improve the scores for forecasts a few days ahead.
(3) Scenario generation. Ten years of weather data were available to the contestants at the beginning of the evaluation period. We expected that some of the contestants would investigate the weather scenario generation methods for probabilistic load forecasting.

One of the challenges that we did not address in this competition is hierarchical probabilistic load forecasting. We avoided this challenge intentionally, for two reasons. Firstly, the probabilistic load forecasting literature is not yet rich, so we wanted this competition to help the community build a solid foundation. Secondly, hierarchical probabilistic load forecasting would require the contestants to submit large datasets, which could have caused technical issues for the newly-developed competition platform.

4.2. Fall 2015 in-class probabilistic load forecasting competition

Tao Hong from UNC Charlotte organized an in-class probabilistic load forecasting competition in Fall 2015, which was also open to external participants. Since this in-class competition was an extended version of GEFCom2014-L, we denote it as GEFCom2014-E in this paper. In total, 16 teams joined the competition, including seven external teams. The setup of this in-class competition was similar to that of GEFCom2014-L. The topic was one-year-ahead probabilistic load forecasting. The competition included five tasks. Six years of hourly temperature data and four years of hourly load data were provided in the first task, then one year of hourly temperature and load data was released incrementally for each of the following

Please cite this article in press as: Hong, T., et al., Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. International Journal of Forecasting (2016), http://dx.doi.org/10.1016/j.ijforecast.2016.02.001

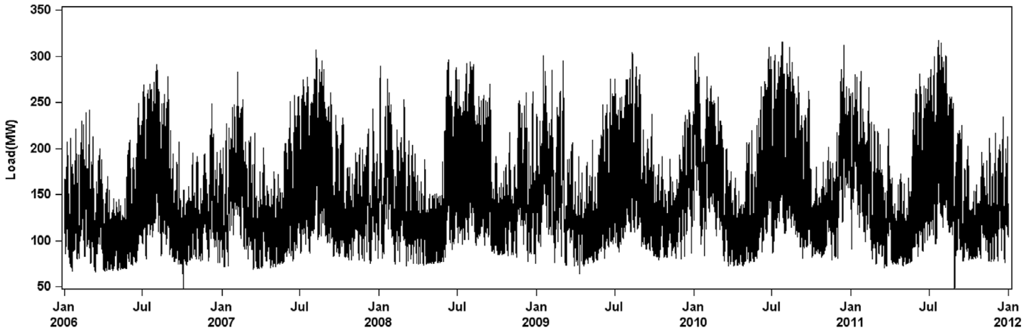

Fig. 2. Load series in GEFCom2014-L.

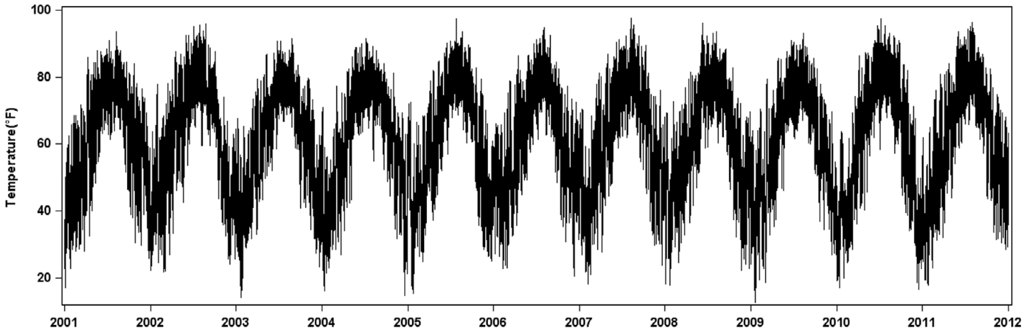

Fig. 3. Temperature series in GEFCom2014-L.

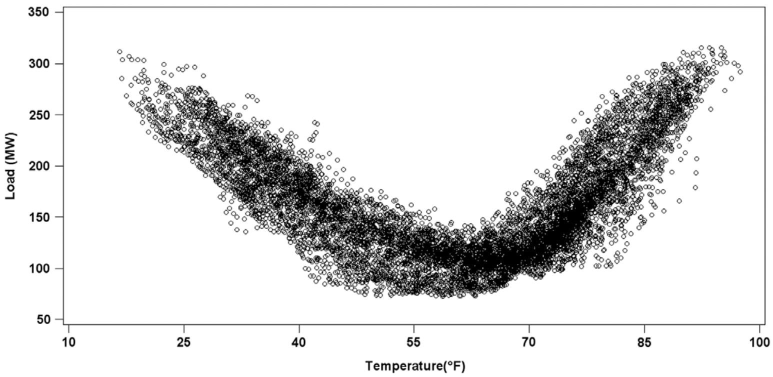

Fig. 4. Scatter plot of the load and temperature in GEFCom2014-L (the year 2010).

four tasks. Figs. 5 and 6 show the load and temperature series respectively. Fig. 7 shows the scatter plot of the load and temperature. The data are published in the online supplements to this paper (see Appendix A).

4.3. Summary of the methods

A total of 16 teams beat the benchmark and submitted final reports in GEFCom2014-L. The top five teams submitted papers to this special issue. In GEFCom2014-E, a total of eight teams were eligible on the final leaderboard. The team Ziel Florian, which took second place, was invited to submit a paper to this special section. Ziel Florian's forecast was compared with the benchmarks developed by Bidong Liu, a top-8 team in GEFCom2014-L. Table 6 summarizes seven aspects of the methods used by the seven teams. The first four are forecasting techniques, data cleansing methods, forecast combination methods and integration methods, which were mentioned as the four main challenges in the introduction section. The last three are the three specific challenges for the load forecasting track, namely weather station selection methods, multi-horizon forecasting methods, and scenario generation methods.

5. Probabilistic electricity price forecasting

5.1. Problem and data description

The aim of GEFCom2014-P was to forecast the probabilistic distribution (in quantiles) of the electricity price for one zone on a rolling basis. The forecast horizon was

Please cite this article in press as: Hong, T., et al., Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. International Journal of Forecasting (2016), http://dx.doi.org/10.1016/j.ijforecast.2016.02.001

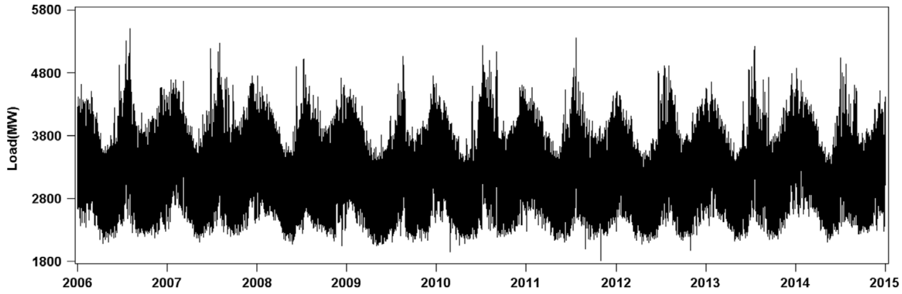

Fig. 5. Load series in GEFCom2014-E.

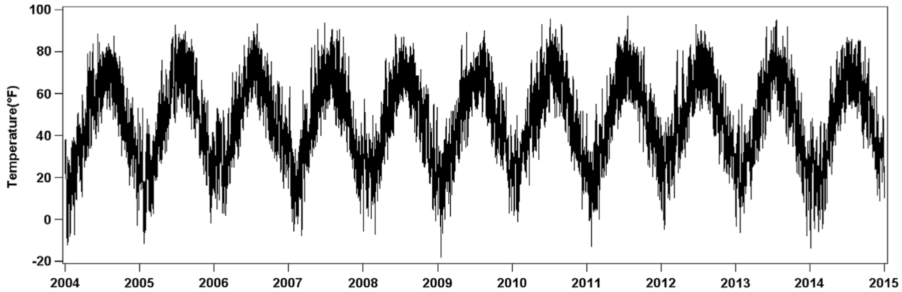

Fig. 6. Temperature series in GEFCom2014-E.

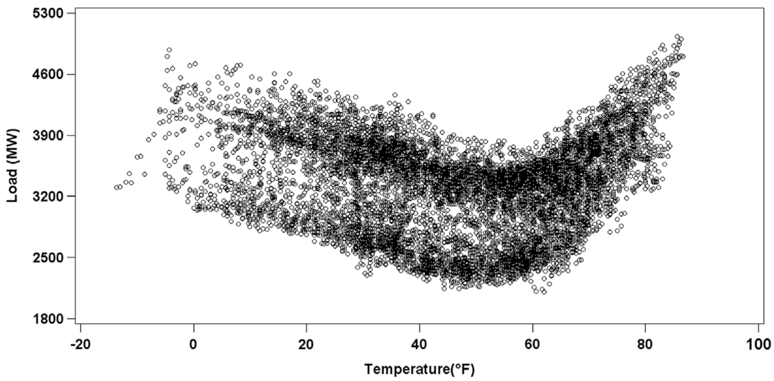

Fig. 7. Scatter plot of the load and temperature in GEFCom2014-E (the year 2014).

24 h. Hourly data were provided, including the locational marginal price, zonal load forecasts and a system load forecast.

The first data release on August 15th included about 2.5 years of hourly prices, and zonal and system load forecast data (from January 1st, 2011, to June 15th, 2013), together with the zonal and system load forecasts for the next day (June 16th, 2013), for which the contestants were asked to forecast the price. Unlike the other three tracks, where the forecast origin was moved forward with each task, the price forecasting track was set up so that the contestants were asked to forecast the following days, as listed in Table 7, using the historical data given right before each forecasted period. In total, the price forecasting track involved about three years of locational marginal price, zonal and system load forecast data (from January 1st, 2011 to December 17th, 2013), as shown in Fig. 8. The data are published along with this paper in the online supplements (see Appendix A).

The main challenge when designing a price forecasting track was to mimic the real-world price forecasting tasks. In reality, the electricity price is driven by other many factors in addition to the load, such as generator outages and transmission congestions. Price forecasters usually perform sophisticated market simulations with all relevant data in order to predict price spikes. However, to keep the competition problem easy for a large data science community to understand, we decided to keep the dataset as simple as possible. Therefore, only zonal and system load forecast data were provided for the forecasting of electricity prices.

5.2. Summary of the methods

A total of 14 teams beat the benchmark and submitted final reports in the probabilistic electricity price forecasting track. The top four teams submitted papers to this special issue. Table 8 summarizes their methods from three

Table 6
Summary of the methods used by the top five teams of the load forecasting track.

Team	Techniques	Data cleansing	Forecast combination	Integration	Weather station selection	Multi-horizon forecasting	Scenario generation
Tololo	Quantile regression; generalized additive models	No	No	Yes. Using cross validation to figure out the cut-off between short and medium forecast horizons.	Four stations (6, 10, 22 and 25) based on Generalized cross validation (GCV).	Yes. One method for forecasting 1 to 48 h ahead, one for forecasting beyond 49 h ahead.	Probabilistic temperature forecasts for medium-term forecasting; 800 temperature scenarios for short-term forecasting.
Adada	Generalized additive models	No	Yes. An exponentially weighted algorithm (EWA).	Yes. Found out weather stations selected for point load forecasting are not optimized for probabilistic forecasting.	Initially seven stations based on GCV. Refined selection using EWA, reaching three stations (6, 10, 13).	No	1000 temperature scenarios from a temperature simulation model.
Jingrui (Rain) Xie	Multiple linear regression, unobserved component models, exponential smoothing models, artificial neural networks, ARIMA, residual simulation	Yes. Model-based outlier detection and data cleansing.	Yes. Averaging point forecasts.	Yes. Optimize the parameters by considering residual simulation with model selection.	11 stations.	No	10 scenarios based on 10 years of temperature history. Also tried shifting the temperature history.
OxMath	Conditional kernel density estimation; quantile regression	No	Yes. Weighted average of the quantile time series for the horizon beyond the first day.	Yes. Through implementation of the hybrid forecast.	No. Using average of all 25 stations.	Yes. Dividing the forecast horizon into five periods.	No
E.S. Mangalova	Nadaraya–Watson kernel regression	No	No	No	No. Using average of all 25 stations.	No	No
Ziel Florian	Lasso; bivariate threshold autoregressive model	No	No	No	Two stations selected based on the goodness-of-fit from a cubic regression.	No	No
Bidong Liu	Multiple linear regression	No	No	No	No. Using average of all 25 stations.	No	No

Table 7
Days to be forecasted in the price forecasting track.

Task #	Forecasted day	Task #	Forecasted day	Task #	Forecasted day
1	June 16th, 2013	6	July 13th, 2013	11	July 24th, 2013
2	June 17th, 2013	7	July 16th, 2013	12	July 25th, 2013
3	June 24th, 2013	8	July 18th, 2013	13	December 7th, 2013
4	July 4th, 2013	9	July 19th, 2013	14	December 8th, 2013
5	July 9th, 2013	10	July 20th, 2013	15	December 17th, 2013

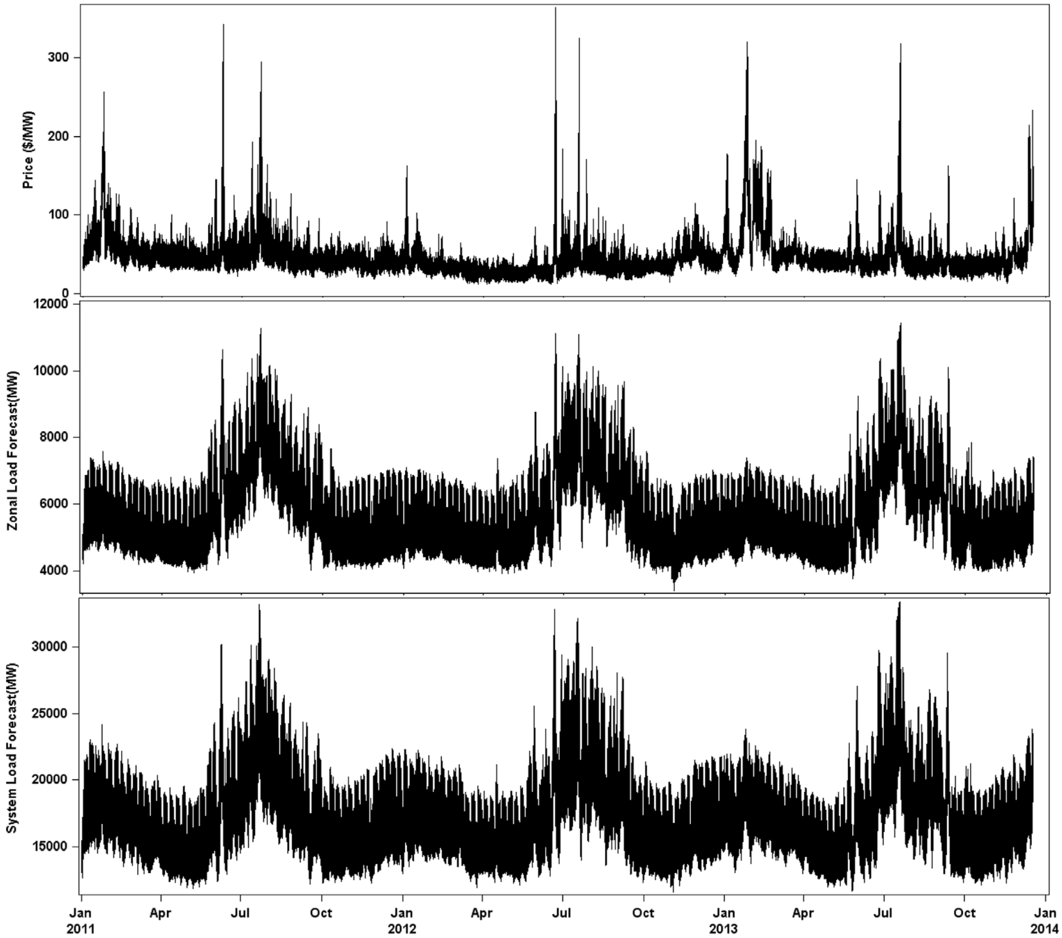

Fig. 8. Locational marginal price series in GEFCom2014-P.

Table 8
Summary of the methods used by the top four teams of the price forecasting track.

	Techniques	Spike preprocessing	Forecast combination
Tololo	(1) Quantile regression, generalized additive models; (2) autoregressive models, random forest regression, gradient boosting machine; (3) Kernel based quantile regression.	Preprocessed spikes for some of the models.	ML-Poly aggregation
Team Poland	Autoregressive models with exogenous variables; filtering; quantile regression; judgmental forecasting	Three filtering methods: day type filtering, similar load profile filtering and expected bias filtering	Arithmetic average
GMD	Feed forward neural network	None	None
C3 Green Team	Quantile regression; radial basis function network; k-means algorithm; alternating direction method of multipliers; Autoregressive models with exogenous variables	None	None

aspects, namely forecasting techniques, price spike preprocessing methods, and forecast combination methods. It is worth noting that three of the top four teams in GEFCom2014-P used quantile regression.

6. Probabilistic wind power forecasting

6.1. Problem and data description

The aim of the probabilistic wind power forecasting track of GEFCom2014 was to predict the wind power generation 24 h ahead in 10 zones, corresponding to 10 wind farms in Australia, on a rolling basis. The wind power output series from these wind farms are shown in Fig. 9. The locations of these 10 wind farms were not disclosed during GEFCom2014. New forecasts were to be issued each day at midnight. Since the period being forecast for each task was one month, each of these 15 tasks required 28–31 24-h forecasts to be issued for each zone. The forecasts were to be expressed in the form of a set of 99 quantiles, with various nominal proportions between 0 and 1.

Please cite this article in press as: Hong, T., et al., Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. International Journal of Forecasting (2016), http://dx.doi.org/10.1016/j.ijforecast.2016.02.001

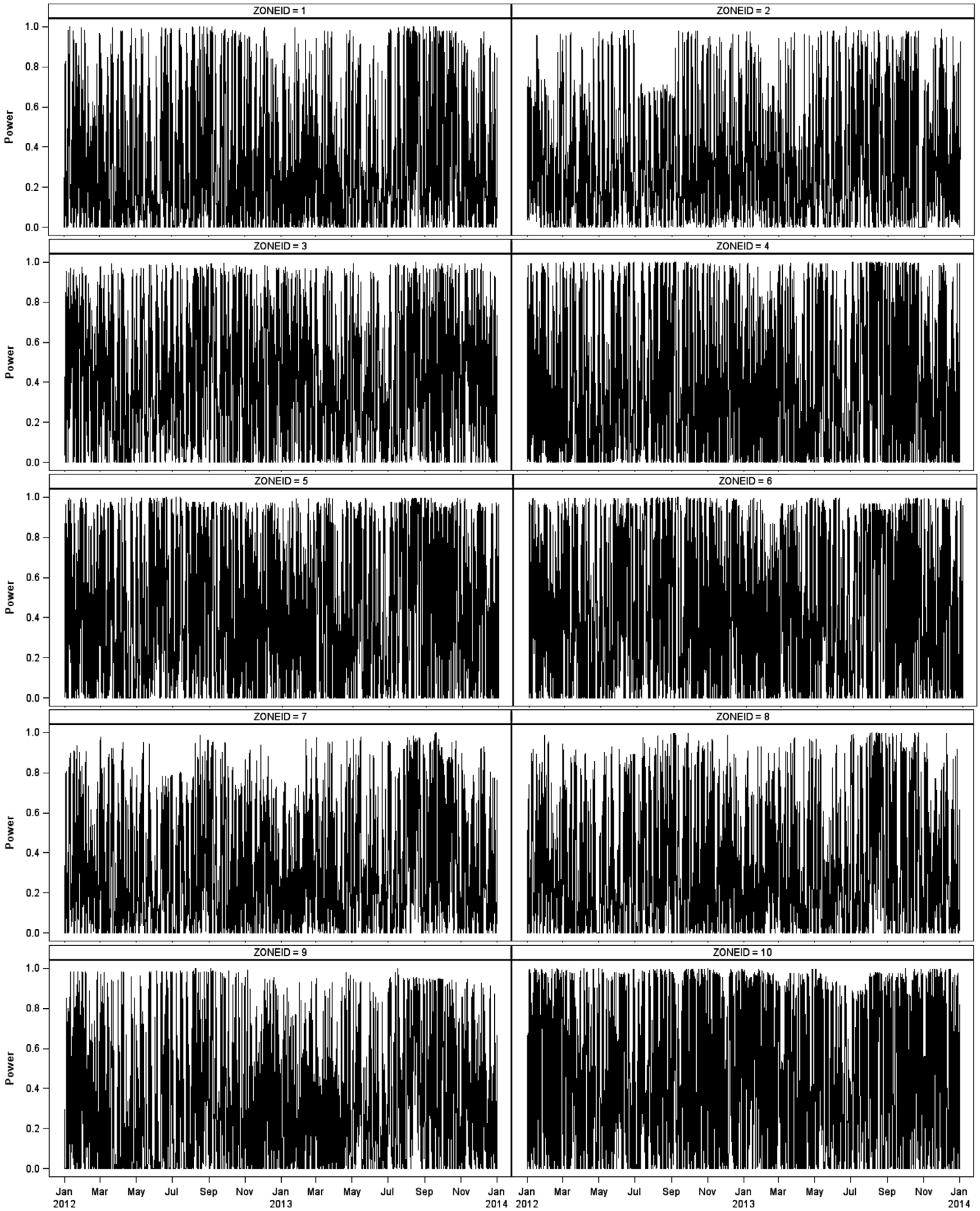

Fig. 9. Zonal wind power generation in GEFCom2014-W.

The predictors included wind forecasts at two heights, 10 and 100 m above ground level, obtained from the European Centre for Medium-range Weather Forecasts (ECMWF). These forecasts were for the zonal and meridional wind components (denoted u and v), i.e., projections of the wind vector on the west-east and south-north axes, respectively. Fig. 10 shows the scatter plots between wind power generation and wind speeds. It was up to the contestants to deduce wind speed and direction forecasts, if necessary. The predictions were provided for the exact locations of these wind farms, issued every day at midnight with an hourly resolution out to 24 h ahead, in line with the specifications of the forecasting exercise. Weather forecasts were available for training, and also as inputs to

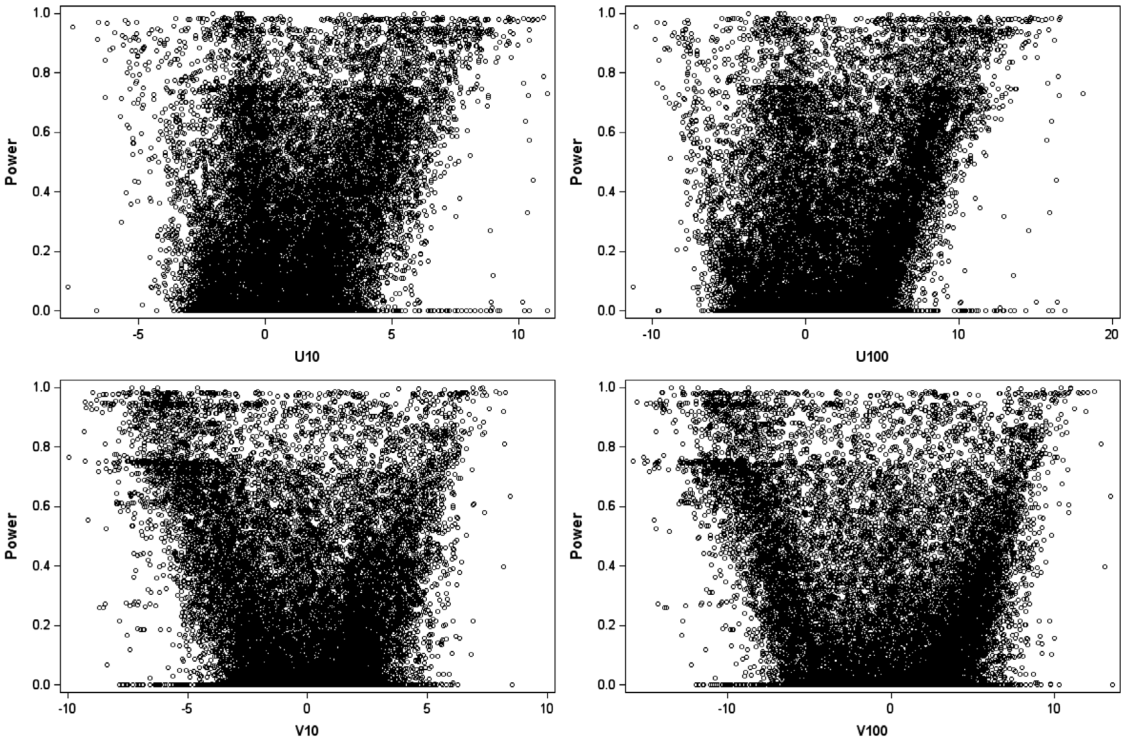

Fig. 10. Scatter matrices of zonal wind power generation and wind speed in GEFCom2014-W.

the various tasks used for forecast evaluation. In addition, power measurements at the various wind farms, with an hourly resolution, were also provided, but only over the training period. All power measurements were normalized by the nominal capacity of the wind farm that they correspond to, while the weather forecasts were left untouched.

Thus, the challenge was to design parametric or nonparametric methods that would allow conditional predictive densities of the wind power generation to be described as a function of the input weather forecasts. Alternative options could have included building a bank of quantile regression models for each location and lead time, based on a wealth of input variables for all wind farms, or more simply formulating a parametric assumption for the predictive densities at each site, with low-order models for their parameters.

6.2. Summary of the methods

The main characteristics of the methods considered by the top-5 teams within this track are summarized in Table 9. They all chose to employ nonparametric approaches to this probabilistic forecasting problem, possibly due to the very definition of the problem, which requested a set of quantiles as a description of predictive densities. Similarly, these teams mostly invested in modern data-mining techniques, such as Gradient Boosting Machines (GBM) and Quantile Regression Forests (QRFs), based on large numbers of input variables and features, possibly considered in multiple layers. Interestingly, some of the teams considered offsite information, e.g., wind forecast data from other wind farms, while others did not. In the frame of GEFCom2012, where the focus was on single-valued predictions and the lead evaluation criterion was the Root Mean Square Error (RMSE), it was clear that employing offsite information was beneficial. Here, though, while the best teams used offsite information, it is not clear whether this was a deciding factor in their outperforming of the others, or whether this was due more to the actual models and forecasting methodologies that they used.

7. Probabilistic solar power forecasting

7.1. Problem and data description

The probabilistic solar power forecasting problem in GEFCom2014 was very similar in design to the wind track described in Section 6. Solar power generation was to be predicted on a rolling basis for 24 h ahead, for three solar power plants located in a certain region of Australia. The solar power generation profiles are shown in Fig. 11. The exact locations of these solar power plants were not disclosed during GEFCom2014. The forecasts were to be issued at midnight each day. Since the forecast period for each task was one month, 28–31 forecast series were issued for each of these 15 tasks. The forecasts were to be expressed in the form of 99 quantiles with various nominal proportions between zero and one.

The data available included weather forecasts for 12 weather variables, as obtained from the European Centre for Medium-range Weather Forecasts (ECMWF). These variables are summarized in Table 10.

The contestants were free to perform variable selection based on all these variables, and/or to generate and

Please cite this article in press as: Hong, T., et al., Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. International Journal of Forecasting (2016), http://dx.doi.org/10.1016/j.ijforecast.2016.02.001

Table 9
Summary of the methods used by the top five teams in the wind track of GEFCom2014.

Team	Parametric/nonparametric	Forecasting models and techniques	Generalization ability (preventing overfitting)	Input variables and features (most important)	Offsite information
kPower	Nonparametric	Gradient Boosting Machines (GBM), organized in two layers	Cross validation	Forecasts of wind speed and direction at 10 and 100 m, at target and neighboring lead times, time of day	Yes
dmlab	Nonparametric	Quantile Regression Forest (QRF) and Gradient Boosting Decision Trees (GBDT)	Cross validation	Forecasts of wind speed and direction at 10 and 100 m, and features derived from these	No
E.S. Mangalova	Nonparametric	k-Nearest Neighbor (k-NN)	Training data selection	Wind components, wind speed and direction at 10 and 100 m (filtered), time of day	Yes
C3 Green Team	Nonparametric	Multiple Quantile Regression (MQR)	Feature selection algorithm, regularization when estimating, and cross-validation	Forecasts of wind speed and direction at 10 and 100 m, and features derived from these, time of day and year	Yes
Yao Zhang	Nonparametric	k-Nearest Neighbor (k-NN) and Kernel Density Estimation (KDE)	Cross-validation	Forecasts of wind speed and direction at 10 and 100 m, time of year, deterministic wind power prediction	No

Table 10
Variables for GEFCom2014-S.

Variable id.	Variable name	Units	Comments
078.128	Total column liquid water (tclw)	kg m^{-2}	Vertical integral of cloud liquid water content
079.128	Total column ice water (tciw)	kg m^{-2}	Vertical integral of cloud ice water content
134.128	Surface pressure (SP)	Pa	
157.128	Relative humidity at 1000 mbar (r)	%	Relative humidity is defined with respect to saturation of the mixed phase, i.e., with respect to saturation over ice below −23 °C and with respect to saturation over water above 0 °C. In the regime in between, a quadratic interpolation is applied.
164.128	Total cloud cover (TCC)	0–1	Total cloud cover derived from model levels using the model's overlap assumption
165.128	10-metre U wind component (10u)	m s^{-1}	
166.128	10-metre V wind component (10v)	m s^{-1}	
167.128	2-metre temperature (2T)	K	
169.128	Surface solar rad down (SSRD)	J m^{-2}	Accumulated field
175.128	Surface thermal rad down (STRD)	J m^{-2}	Accumulated field
178.128	Top net solar rad (TSR)	J m^{-2}	Net solar radiation at the top of the atmosphere. Accumulated field
228.128	Total precipitation (TP)	m	Convective precipitation + stratiform precipitation (CP + LSP). Accumulated field.

select new features. The predictions were provided for the exact locations of the solar power plants, and were issued every day at midnight with an hourly resolution out to 24 h ahead, in line with the specifications of the forecast exercise. Weather forecasts were available for training, and also as inputs to the forecasting exercises. Power measurements with an hourly resolution were also provided, but only over the training period.

7.2. Summary of the methods

Although the setup for the forecast competition was very similar to that for wind power, the solar power case was a bit more subtle, in view of the larger number of input variables, and also since there are additional factors (time of day, clear sky, etc.) that may affect solar power generation and its dynamics. A summary of the features of the approaches employed by the five best teams in this track is given in Table 11.

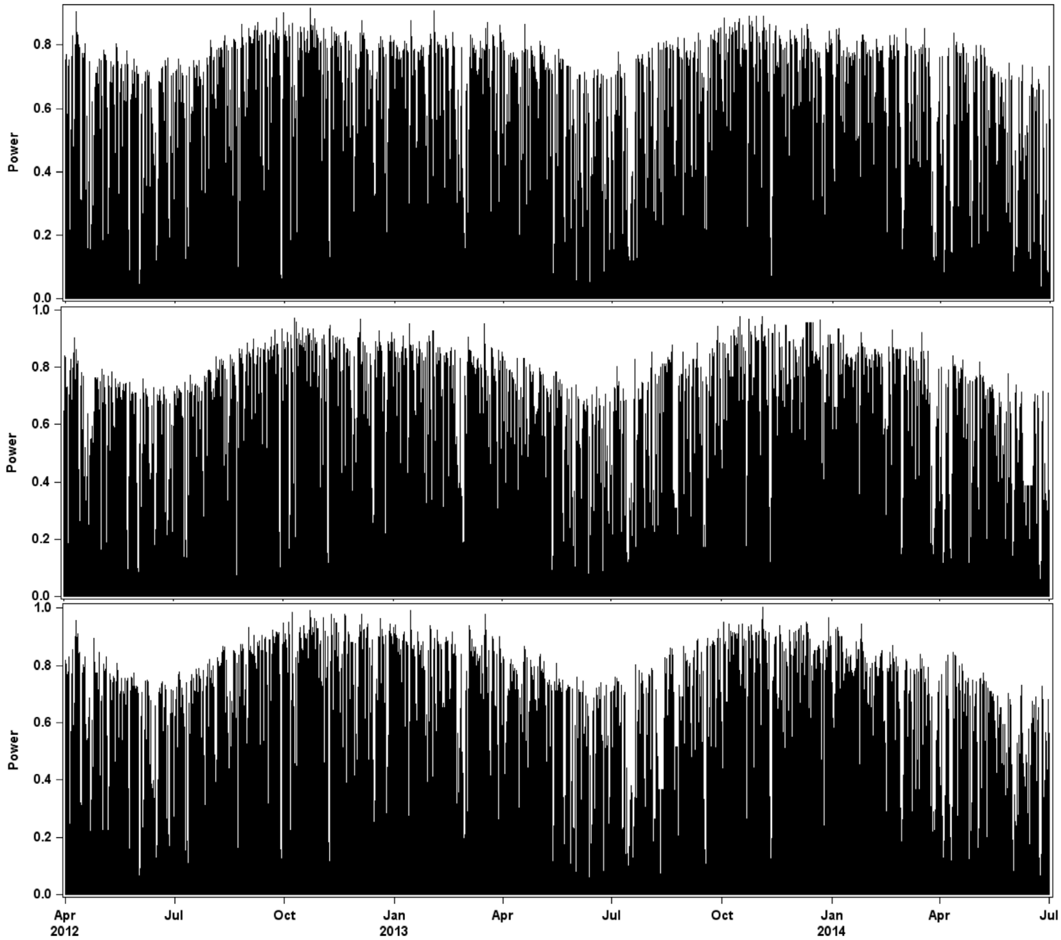

Fig. 11. Zonal solar power generation in GEFCom2014-S.

All of the approaches that ranked highly in this competition were nonparametric, and most used variants of quantile regression and gradient boosting. It should be noted that, as for the wind power case, the team that ranked second did not use offsite information, while all of the others did. This could mean that, while it is intuitive that offsite information could help to improve the forecast quality, large gains may be possible simply by carefully designing a forecasting approach that accounts for local information only. Considering offsite information increases dimensionality substantially, making the variable/feature selection more complex. In the present case, all of the best teams had methodologically sound strategies for defining the variables and features that were used as inputs for their forecasting. They were actually very creative in deriving new features from the variables originally provided, using differentiation from accumulated fields, time shifting, integration, etc.

8. A 10-year ahead forecast of energy forecasting

8.1. The research maturity quadrant

Fig. 12 depicts the maturity level of each subdomain (demand, price and supply) of energy forecasting, for point and probabilistic forecasting. We separated long term (a few months to a few decades ahead) and short term (two weeks ahead or shorter) load forecasting in the area of demand forecasting, due to their distinct characteristics. The temperature is a key factor driving the load. Because the temperature is fairly predictable in the short term, we can generate fairly accurate short term load forecasts. Many statistical and artificial intelligence techniques have been tried for short term point load forecasting in the literature, which makes STLF the most mature for point forecasting. Since the temperature is less predictable in the long term, point load forecasts are considerably less skillful. On the other hand, researchers have presented several appealing case studies for long term probabilistic load forecasting, bringing LTLF second highest for probabilistic forecasting maturity.

Wind power forecasting takes the highest place in probabilistic forecasting maturity. This is largely because wind power forecasting is the closest to meteorological forecasting, where probabilistic forecasting is well-established and commonly accepted. Although solar power forecasting is in the same subdomain (renewable generation) as wind power forecasting, it is not as mature as WPF for either point or probabilistic forecasting. This is due to the fact that solar power penetration has not been significant enough

Table 11
Summary of the methods used by the top five teams in the solar track of GEFCom2014.

Team	Parametric/nonparametric	Forecasting models and techniques	Generalization ability (preventing overfitting)	Input variables and features (most important)	Offsite information
Gang-gang	Nonparametric	Gradient Boosting (GB) and *k*-Nearest Neighbour (*k*-NN)	Cross-validation	Clear sky model, as well as all variables provided	Yes
dmlab	Nonparametric	Quantile Regression Forest (QRF) and Gradient Boosting Decision Trees (GBDT)	Cross validation	Variables provided, time of day and of year, differentiated variables (for the accumulated fields)	No
C3 Green Team	Nonparametric	Multiple Quantile Regression (MQR)	Feature selection algorithm, regularization when estimating, and cross-validation	Wealth of features based on all input variables, time of day and time of year	Yes
Giuseppe Casalicchio	Nonparametric	Quantile Regression and Quantile Regression Forest (QRF)	Lasso penalization	Wealth of features based on all input variables, considering lagging, smoothing, and combination	Yes
UT_Argonne	Nonparametric	Ensemble of Random Forest (RF), Gradient Boosting Machines (GBM) and Support Vector Machines (SVM)	Training data selection	Wealth of features based on all input variables, considering time shifting, integration, etc.	Yes

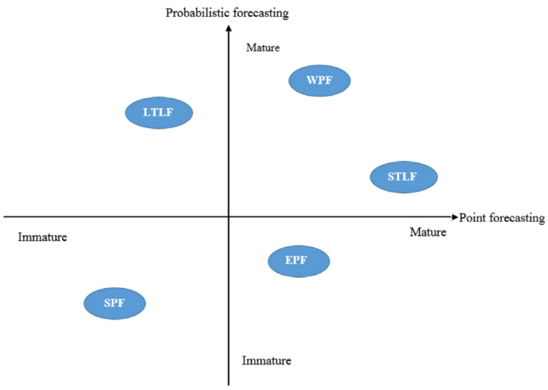

Fig. 12. Maturity quadrant of the energy forecasting subdomains (SPF: solar power forecasting; LTLF: long term load forecasting; EPF: electricity price forecasting; WPF: wind power forecasting; STLF: short term load forecasting).

until the last few years. This is why we separated wind and solar power forecasting in Fig. 12.

Research into price forecasting took off right after the deregulation of the electric power industry. Most papers in the literature take a data-driven approach, without modeling the market or the circuit. Unlike wind power, which is influenced chiefly by one or two factors, namely the wind speed and direction, electricity prices, and especially price spikes, are influenced heavily by a wide range of factors other than the electricity demand, such as transmission congestion, generation outage, market participant behaviors, etc. These factors, and the uncertainties associated with them, are hard to incorporate into statistical or artificial intelligence models. Therefore, the point forecasting maturity of prices is less than that of wind power or short term loads. The probabilistic forecasting maturity of prices is also less than that of either wind or loads.

8.2. Twelve predictions for the next decade of energy forecasting

To conclude this paper, we make a 10-year ahead forecast for the field of energy forecasting, which is summarized into the following 12 predictions:

(1) Solar power forecasting research will flourish
With the increasing penetration of solar power, in the forms of both rooftops and solar farms, the research progress of solar power forecasting is expected to advance greatly over the next decade.
(2) The development of practical error measures for probabilistic energy forecasting
The probabilistic wind power forecasting literature has long embraced proper scoring rules for evaluating probabilistic forecasts. The GEFCom2014 has now brought the pinball loss function to the attention of the wider energy forecasting community. Despite its simplicity from an academic standpoint, the pinball loss function has still been difficult for industry practitioners to adopt. The proposal and widespread use of more practical error measures, similar to the way in which the MAPE is often used as a reference for load forecasting, or the RMSE for wind power forecasting, is expected to fill this gap.
(3) A connection between probabilistic and point energy forecasting

One observation from Fig. 12 is that, for most of these subdomains, the maturity of point forecasting is not aligned with that of probabilistic forecasting. This is due primarily to an important but unanswered question: can a better point forecasting model lead to a better probabilistic forecast? If such is the case, we would be able to take advantage of the extensive point energy forecasting literature to enhance probabilistic energy forecasts across all subdomains.

(4) An increased use of high resolution data, temporally, spatially and conceptually
Modern information technologies have enabled energy companies and weather service providers to make measurements at very high resolutions, which can go sub-second, within acres, and for a large variety of meteorological variables. We expect to see such rich information sets being used more widely by the energy forecasting community.

(5) The unification of energy forecasting methodologies
Several teams in GEFCom2014 won multiple tracks using similar methodologies. This fact indicates that there are some general methodologies or frameworks that can be used across a range of subdomains of probabilistic energy forecasting. Essentially, probabilistic energy forecasting is a branch of probabilistic forecasting. Over the next decade, it is expected that energy forecasters will explore the probabilistic forecasting literature further, and adopt the most effective methods for probabilistic energy forecasting.

(6) A diversification of energy forecasting subjects
The evolution of the electric grid has provided many emerging problems for the energy forecasting community to deal with, such as forecasting the demand at the premises level, forecasting the trend of rooftop PV penetration, forecasting changes in load due to demand response programs, and forecasting system interruptions due to severe weather conditions. Thus, it is expected that the energy forecasting community will have to tackle many more problems like these in the near future.

(7) The fusion of energy forecasting problems
Most energy forecasting papers in the literature focus on specific subdomains. More and more cross-subdomain energy forecasting papers are expected to appear in the literature, due to the fusion of energy forecasting problems. For instance, when residential customers install solar rooftops without sub-metering, the traditional load forecasting and solar power forecasting problems fuse into net demand forecasting. Similarly, prices are driven by both supply and demand. Load control activities, together with an increased penetration of renewable energy, are fusing the demand, supply and price forecasting problems.

(8) Interdisciplinary collaborations with other communities
The maturity of probabilistic wind power forecasting is due largely to its collaboration with the meteorological forecasting community; one weakness of electricity price forecasting is the difficulty of including market and circuit models; and the advancement of short term load forecasting is due partly to the involvement of both the statistics and artificial intelligence communities. These past examples have shown us the importance of interdisciplinary collaborations for the advancement of energy forecasting research progress. As various communities become more involved in energy forecasting, our knowledge will be advanced further.

(9) Additional energy forecasting competitions
Over the past two decades, forecasting competitions, from the EPRI and EUNITE competitions in the 1990s to the GEFCom2012 and GEFCom2014 in the 2010s, have been playing an important role in encouraging research progress and attracting data scientists across a range of disciplines. It is therefore essential to keep organizing new energy forecasting competitions.

(10) Regular conferences in energy forecasting
There are not yet regular conferences dedicated to energy forecasting. Over the past few years, energy forecasting has been part of the technical programs of several large conferences, such as the IEEE Power and Energy Society General Meeting, the International Symposium on Forecasting (ISF), and the International Conference on Energy & Meteorology (ICEM). The International Symposium on Energy Analytics has been scheduled right before ISF2017 in Australia. In future, we hope to make this a regular event.

(11) A dedicated publication outlet for energy forecasters
There is not yet a journal dedicated to energy forecasting either. Top quality energy forecasting papers have been published in journals such as the *International Journal of Forecasting*, *IEEE Transactions on Smart Grid*, *IEEE Transactions on Power Systems*, *IEEE Transactions on Sustainable Energy*, *Solar Energy*, and *Wind Energy*. In recent years, the community has been publishing hundreds of energy forecasting papers each year. We believe that there will soon be enough interest to support an energy forecasting journal.

(12) A society for energy forecasters
The researchers working on energy forecasting are spread across various different technical societies, such as the IEEE Power and Energy Society, the International Institute of Forecasters, and the American Meteorological Society. Some of the industry practitioners have also been gathering at local society meetings, such as the Association of Edison Illuminating Companies Load Research Conference and the Edison Electric Institute Load Forecasting Group Meeting. We believe that the next decade will be a good time to bring researchers and practitioners together to form a society for energy forecasters.

Acknowledgments

We would like to acknowledge the careful review performed by Rafal Weron, whose comments helped further improve the clarity of this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.ijforecast.2016.02.001.

References

Bacher, P., Madsen, H., & Nielsen, H. A. (2009). Online short-term solar power forecasting. *Solar Energy*, *83*(10), 1772–1783.

Gneiting, T., & Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, *1*, 125–151.

Hong, T. (2014). Energy forecasting: past, present, and future. *Foresight: The International Journal of Forecasting*, *32*, 43–49.

Hong, T., & Fan, S. (2016). Probabilistic electric load forecasting: a tutorial review. *International Journal of Forecasting*, http://dx.doi.org/10.1016/j.ijforecast.2015.11.011. in this issue.

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

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

Hyndman, R. J., & Fan, S. (2010). Density forecasting for long-term peak electricity demand. *IEEE Transactions on Power Systems*, *25*(2), 1142–1153.

Lorenz, E., Scheidsteger, T., Hurka, J., Heinemann, D., & Kurz, C. (2011). Regional PV power prediction for improved grid integration. *Progress in Photovoltaics: Research and Applications*, *19*(7), 757–771.

Pinson, P. (2013). Wind energy: forecasting challenges for its operational management. *Statistical Science*, *28*(4), 564–585.

Weron, R. (2014). Electricity price forecasting: a review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, *30*(4), 1030–1081.

Zhang, Y., Wang, J., & Wang, X. (2014). Review on probabilistic forecasting of wind power generation. *Renewable and Sustainable Energy Reviews*, *32*, 255–270.

Please cite this article in press as: Hong, T., et al., Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. International Journal of Forecasting (2016), http://dx.doi.org/10.1016/j.ijforecast.2016.02.001