HSC Research Report

HSC/13/16

Global Energy Forecasting Competition 2012

Tao Hong[1]
Pierre Pinson[2]
Shu Fan[3]

[1] SAS Institute Inc, United States
[2] Technical University of Denmark, Denmark
[3] Monash University, Australia

Hugo Steinhaus Center
Wrocław University of Technology
Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland
http://www.im.pwr.wroc.pl/~hugo/

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.

Key words: energy forecasting; forecasting competition; load forecasting; wind power forecasting.

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. //实用产业

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//标杆 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//最先进的 techniques for energy forecasting, (iii) bridge the gap between academic research and industry practice,//行业惯例 (iv) promote analytics in power and energy education, and (v) prepare the industry to overcome the forecasting challenges posed by the smart grid world.//智能电网的世界 The competition included two tracks, hierarchical load forecasting //层次化负荷预测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

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

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:

$$\sqrt{\frac{\sum_i \quad _i(A \qquad _i)^2}{\sum_i}},$$

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 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 station separately, and the best three were used for each zone	No	Averaging the temperatures during the same days from previous years	No
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

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

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//容量 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. //与电力系统运营和参与电力市场有关的一些决策问题需要投入某种形式的预测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//简单时间序列模型 for wind forecasting at a site of interest, then converted the resulting wind forecasts to electric power generation//发电 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//新千年, 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 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. //数据清理、集成预测和一体化方面

3.2. Data description

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.

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. //数据信息

//它们通过风电场各自的标称容量进行标准化，以获得0到1之间的标准化功率值，从而使风电场的原始特性得以掩盖。这也使得对各种风电场的预测结果进行无标度比较成为可能。

//在地面以上10米处收集了地面风的纬向（u）和经向（v）分量的气象预报。它们摘自欧洲中期天气预报中心（ECMWF）的档案。ECMWF每天在00UTC和12UTC发布两次高分辨率确定性预测，时间分辨率在3小时到10天之间。为了匹配功率测量的小时分辨率（这也是大多数预测应用所要求的），使用三次样条插值预测，以便具有小时分辨率。数据集中只整理了每个预测系列的前48小时。请注意，这些气象预报也以风速和风向的形式提供给那些更愿意使用这种格式的人。

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 //11个电子表格 (in comma-separated values CSV format) for the wind power forecasting track//逗号分隔值文件格式CSV
如果你的机器上装了 Microsoft Excel的话，.csv 文件默认是被Excel打开的。需要注意的是，当你双击一个.CSV 文件，Excel 打开它以后即使不做任何的修改，在关闭的时候 Excel 往往会提示是否要改成正确的文件格式，这个时候如果选择“是”，因为 Excel 认为.CSV 文件中的数字是要用科学记数法来表示的，Excel 会把 CSV 文件中所有的数字用科学计数来表示（2.54932E+5 这种形式），这样操作之后，只是在 Excel 中显示的时候会不正常，而 csv 文件由于是纯文本文件，在使用上没有影响；如果选择了“否”，那么会提示你以 xls 格式另存为 Excel 的一个副本。

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.

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

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.
3.3条。方法和结果总结

持久性方法，作为发布这些风电场风力发电量预测的最简单方法之一，在这里被用作基准。这种预测方法基于随机游走模型，其中预测值被定义为最新的可用观测值。表4总结了九个选定小组共同使用的方法。我们还显示了这9个团队的错误统计数据和表5中的持久性基准，以及每个团队提交的数据。错误统计（在RMSE中）按风电场、验证数据、测试数据和所有数据进行细分。

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

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

Kaggle ID	Technique	Data Cleansing	Ensemble Forecasting
Leustagos	九个模型的线性组合（从气象预报到电力的回归、风电场之间的依赖关系、自回归组件，具有不同的模型结构）	No	Yes
DuckTile	数据清理，然后以风预报、一年中的日期和时间为输入进行局部线性回归	Yes	No
MZ	基于径向基函数和自回归特征的正则化最小二乘估计线性模型	No	No
propeller	从风预报到功率测量的线性回归，然后用梯度助推机进行非线性校正（通过交叉验证确定最佳输入）	Yes	No
Duehee Lee	大量神经网络（52）和高斯过程模型（5）的简单组合，将所有输入数据映射到功率测量	No	Yes
MTU EE5260	气象预报功率转换的线性回归和神经网络方法	No	No
SunWind	功率曲线模型、自回归模型、局部线性回归模型和支持向量机模型的简单组合	No	Yes
ymzsmsd	所有风电场输入测量和预测的稀疏贝叶斯学习	No	No
4138 Kalchas	基于正则核心回归的气象预报功率转换	No	No
Benchmark	坚持不懈	No	No

Table 5. Error statistics (RMSEs) of selected entries in the wind power forecasting track.风电预测轨道中选定项目的误差统计（RMSE）

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

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

4. Discussion

图1显示了从比赛开始时从每个赛道下载数据的唯一ID的累计数量。垂直的虚线表示比赛结束，在这一点上，每个赛道有大约600个唯一的ID。比赛结束后，Kaggle用户仍在下载这些数据。利用Kaggle的平台，这次竞赛吸引了比预期更多的参与者，其中许多人都是公用事业行业以外非常有经验的数据科学家。虽然参赛者背景的多样性给能源预测领域带来了很多新的思路，但也有一些参赛者对参加赛后活动不感兴趣，比如提交报告、在会议上介绍自己的工作、撰写科学论文等。

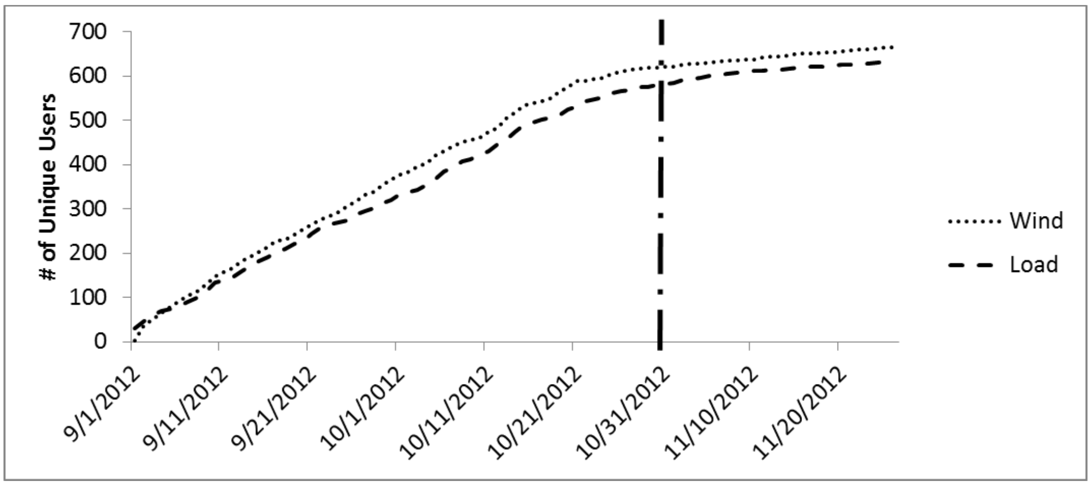

图1。在前3个月内从两个赛道下载比赛数据的唯一用户数。

Kaggle提供了一个论坛，参与者和竞争管理者可以在其中发布问题、答案和调查结果。此功能允许参与者在公共域中互相帮助。它还允许管理员在提出问题时立即解决问题。随着竞争的进行，论坛中有丰富的信息，这就要求新的参与者对旧的帖子进行评论。一些参与者，主要是新的Kaggle用户，可能不会查看以前的帖子，这可能导致违反一些竞争规则。为了避免今后出现类似情况，我们建议竞争管理人员提高与会者对论坛讨论中重要职位的认识。

在分层负荷预测轨道中，为了保持各区域的负荷水平，我们给出了实际负荷而不是标准化的负荷值，这就使得一些参与者能够猜测电力公司的位置，并利用外部信息赢得竞争。为了避免这种情况，我们要求团队提交报告和代码，然后由GEFCom2012奖励委员会进行评估。最后，两个小组由于使用了预测周的实际温度数据而被取消资格。在风力发电预测轨道中，数据是标准化的，因此参与者无法通过猜测风电场的位置来找到解决方案。

在现实世界的短期负荷或风力预测工作中，预报员必须利用最新的可用信息，每天进行预报。换句话说，预测原点每天都在移动。In order to implement this feature in a competition, we would have to host multiple phases, with new data being released at each phase. 虽然每个阶段可能需要几个星期才能完成，但整个比赛将需要两个月以上的时间。实现这一特点还需要参赛者在整个比赛中充分参与。作为一个班内比赛，这比首届国际比赛更容易实现，因此，我们在设计GEFCom2012时没有设置此功能。作为一个修正案，我们在历史上留下了一些缺失的时期以供预测。由于我们无法确定参与者在预测缺失时段时是否使用缺失时段之后的数据，因此我们没有限制参与者仅使用预测每个缺失时段之前的数据。这种设置可能意味着回归或其他一些数据挖掘技术比某些时间序列预测技术（如ARIM A）具有优势，这可能是我们在分层负荷预测轨道中没有使用Box-Jenkins方法收到任何报告的部分原因。

从本质上讲，预测是一个随机问题。在公用事业行业中，一些公用事业的应用需要以预测密度或情景形式作为输入的概率预测，例如系统规划的年度峰值需求预测（Hynd man&Fan，2012）、系统储备量化（M atos&Bessa，2011年）、机组承诺（Tuohy、M eibom、Denny和O’M alley，2009年）和风力发电交易（Pinson、Chevallier 和Kar iniotakis，2007年）。另一方面，许多决策过程都只进行点预测。大多数能源预测文献都考虑了点预测。在GEFCom2012中，为了使竞争问题和错误分数保持简单明了，我们让参与者开发点预测，而不是概率预测。

5. Conclusion

GEFCom2012包括两个轨道：分级负荷预测和风电预测。比赛吸引了全世界数百人参加。本文从背景、问题、数据、方法、结果和经验教训等方面介绍了GEFCom2012。我们还公布了每个轨道的完整数据集，试图为能源预测建立一个数据池。在未来，我们希望通过增加更多的轨道来扩大竞争，例如长期负荷预测、价格预测和太阳能发电预测。我们还想探索其他特性，例如滚动预测的来源、综合误差分数和概率预测。

Acknowledgement

作者感谢以下组织的成功：IEEE电力与能源协会，IEEE电力系统规划与系统实施委员会，IEEE电力与能源教育委员会，IEEE能源预测工作组，Kaggle，WeatherBank公司，国际预测杂志，以及智能电网的IEEE交易。作者还感谢以下人员对组织GEFCom2012的支持：David Hamilton、Eric Wang、Hamid eza Zareipour、ML Chan、Rob J Hyndman、Wei Jen Lee、Fangxing Li、Shanshan Liu、Anil Pahwa、Mohammad Shahidehpour和Kumar Venayagamothy。

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

参考文献

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 //个人简历

陶红在北京清华大学获得自动化学士学位，在北卡罗莱纳州立大学获得电气工程硕士学位、运筹学和工业工程硕士学位以及运筹学和电气工程博士学位。他是SAS Institute Inc.的能源预测主管、IEEE能源预测工作组主席、全球能源预测竞赛的总主席，以及IEEETransactions on on Smart Grid关于能源预测分析和智能电网应用的特刊的客座主编。他的研究兴趣是分析和实用程序。

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 (HUS~~T), Wuhan, China, in 1995, 2000, an~~d ~~2004, respectively. He conducted pos~~tdoctoral 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.

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

HSC Research Report Series 2013

For a complete list please visit http://ideas.repec.org/s/wuu/wpaper.html

01 *Forecasting of daily electricity spot prices by incorporating intra-day relationships: Evidence form the UK power market* by Katarzyna Maciejowska and Rafał Weron
02 *Modeling and forecasting of the long-term seasonal component of the EEX and Nord Pool spot prices* by Jakub Nowotarski, Jakub Tomczyk and Rafał Weron
03 *A review of optimization methods for evaluation of placement of distributed generation into distribution networks* by Anna Kowalska-Pyzalska
04 *Diffusion of innovation within an agent-based model: Spinsons, independence and advertising* by Piotr Przybyła, Katarzyna Sznajd-Weron and Rafał Weron
05 *Going green: Agent-based modeling of the diffusion of dynamic electricity tariffs* by Anna Kowalska-Pyzalska, Katarzyna Maciejowska, Katarzyna Sznajd-Weron and Rafał Weron
06 *Relationship between spot and futures prices in electricity markets: Pitfalls of regression analysis* by Michał Zator
07 *An empirical comparison of alternate schemes for combining electricity spot price forecasts* by Jakub Nowotarski, Eran Raviv, Stefan Trueck and Rafał Weron
08 *Revisiting the relationship between spot and futures prices in the Nord Pool electricity market* by Rafał Weron and Michał Zator
09 *Rewiring the network. What helps an innovation to diffuse?* by Katarzyna Sznajd-Weron, Janusz Szwabiński, Rafał Weron and Tomasz Weron
10 *Going green: Agent-based modeling of the diffusion of dynamic electricity tariffs* by Anna Kowalska-Pyzalska, Katarzyna Maciejowska, Katarzyna Sznajd-Weron, Karol Suszczyński and Rafał Weron
11 *Forecasting of daily electricity prices with factor models: Utilizing intra-day and inter-zone relationships* by Katarzyna Maciejowska and Rafał Weron
12 *Computing electricity spot price prediction intervals using quantile regression and forecast averaging* by Jakub Nowotarski and Rafał Weron
13 *Long term probabilistic load forecasting and normalization with hourly information* by Tao Hong, Jason Wilson and Jingrui Xie
14 *Fuzzy interaction regression for short term load forecasting* by Tao Hong and Pu Wang
15 *Energy forecasting: Past, present and future* by Tao Hong
16 *Global Energy Forecasting Competition 2012* by Tao Hong, Pierre Pinson and Shu Fan