

Description

Background

In today's competitive and dynamic environment, more and more decision making processes in the power and energy industry are relying on probabilistic forecasts. The applications of probabilistic energy forecasts spread across planning and operations of the entire energy value chain. The Global Energy Forecasting Competition 2014 (GEFCom2014) will bring together state-of-the-art techniques and methodologies for probabilistic energy forecasting. GEFCom2014 features four tracks: Probabilistic Electric Load Forecasting, Probabilistic Electricity Price Forecasting, Probabilistic Wind Power Forecasting and Probabilistic Solar Power Forecasting. For more background reading materials, please visit <http://blog.drhongtao.com/2014/08/recommended-papers-for-gefcom2014-contestants.html>.

Objective

The topic of the probabilistic solar power forecasting track is to forecast the probabilistic distribution (in quantiles) of the solar power generation for 3 solar farms on a rolling basis. The three solar farms are adjacent. Their installation parameters are:

Zone1: Altitude (595m) Panel type(Solarfun SF160-24-1M195) Panel number (8) Nominal power (1560W) Panel Orientation (38° Clockwise from North) Panel Tilt (36°)

Zone2: Altitude(602m) Panel type(Suntech STP190S-24/Ad+) Panel number (26) Nominal power (4940W) Panel Orientation (327° Clockwise from North) Panel Tilt (35°)

Zone3: Altitude(951m) Panel type(Suntech STP200-18/ud) Panel number (20) Nominal power (4000W) Panel Orientation (31° Clockwise from North) Panel Tilt (21°)

Data Description

The target variable is solar power. There are 12 independent variables from the ECMWF NWP output to be used as below, and they are standardized/normalized.

078.128 Total column liquid water (tclw) - unit (kg m^{-2}) - Vertical integral of cloud liquid water content

079.128 Total column ice water (tcIW) - unit (kg m^{-2}) - Vertical integral of cloud ice water content

134.128 surface pressure (SP) - Unit: Pa

157.128 Relative humidity at 1000 mbar (r) -unit (%) - Relative humidity is defined with respect to saturation of the mixed phase, i.e. with respect to saturation over ice below -23C and with respect to saturation over water above 0C. In the regime in between a quadratic interpolation is applied.

164.128 total cloud cover (TCC) - Unit: (0-1) - Total cloud cover derived from model levels using the model's overlap assumption

165.128 10 metre U wind component (10u) - unit (m s⁻¹)

166.128 10 metre V wind component (10V) - unit (m s⁻¹)

167.128 2 metre temperature (2T) - Unit: K

169.128 surface solar rad down (SSRD) - Unit: J m⁻² - Accumulated field

175.128 surface thermal rad down (STRD) - Unit: J m⁻² - Accumulated field

178.128 top net solar rad (TSR) – Unit: J m⁻² - Net solar radiation at the top of the atmosphere. Accumulated field

228.128 total precipitation (TP) - Unit: m - Convective precipitation + stratiform precipitation (CP +LSP). Accumulated field.

Solvers Expectations

During the competition period, the contestants are required to submit probabilistic forecasts in the form of quantiles following the exact format provided in the template file.

Evaluation Metric

Pin ball function will be used to evaluate the accuracy of probabilistic forecasts. The winning teams will be determined by the prize committee based on the forecasting accuracy and final reports. The winning institute will be determined based on the performance of the participating teams.

Contest Timeline

Important Dates:

15 Aug. 2014	Historical data release (competition starts)
14 Sept. 2014	Evaluation period starts
10 Oct. 2014	Registration deadline
6 Dec. 2014	Evaluation period ends
15 Dec. 2014	Final report and code due (competition ends)

Prizes

Team prize:

- 1st place - \$2500
- 2nd place - \$500
- 3rd place - \$500

Guidelines

Data Guidelines:

There are three data files available for each week of the competition period. The predictor file ("predictorsXX.csv") contains all the independent variables during both the training period and the test period. The training file ("trainXX.csv") contains the target variable during only the training period. And the template file ("templateXX.csv") provides a template of the final output. The TIMESTAMP is in UTC time zone

The benchmark in the solar forecasting track was created by taking the same month last year as the predicted solar power across all quantiles. This is a naïve benchmark, which takes a naïve forecast and expands it to 99 quantiles. The contestants do NOT have to provide the same value across all quantiles.

General Guidelines:

Data can be download from "Data" tab. Each task has its own data files and submission templates. Please use the data files provided for each task. Each task will have its own section under "Data" tab, which will be releaved when the task starts. All tasks have been mentioned under "Tasks" tab with their respective start and end dates.

Submission Guidelines:

The submissions for leaderboard will be ONLY be accepted in the template file provided in CSV format for each task. Please see "Make a Submission" page for more details and to download template file. The system will not process the submissions if this format is not followed, this is a strict guideline. Scores will be processed automatically by system. Please refresh the page to see the score. Scores calculation will usually take 1-5 mintues. Solvers can view their past task submissions via this page by following the links for each task.

Please note the following special instructions for GEF Competition:

1. The participants are required to submit their contact information prior to the registration close date to be eligible for the positions on the leaderboard.
2. The number of submissions is limited at 20 entries per pay during the trial period (i.e. Trial Tasks under Tasks Tab) and **1 entry per day during the evaluation period** (i.e. Interim Tasks under Tasks Tab). When determining the winners, preference will be given to those who made minimal number of submissions during the evaluation period.
3. Team size is up to 4 people.
4. Team mergers are not allowed.
5. One account per participant.
6. The competition organizers have the right to publicly disseminate any entries.
7. Winning solutions need to be made available under a popular OSI-approved license in order to be eligible for recognition and prize money.
8. The participants should submit their algorithms, models, complete code and a detailed report via email to the General Chair of GECom2014 (Dr. Tao Hong, hongtao01@gmail.com) and copy the corresponding Vice Chair before the competition end date.

9. Models should be generated using software that is free and open source or is in the following list of exceptions: MATLAB, SAS, R, MS Excel. The participants who are using other software packages should send a free evaluation copy of the software to GEFCom Executive Committee to reproduce and evaluate the proposed models.

The participating teams are required to present the methodologies at IEEE PES General Meeting 2015 to be eligible for the prize.

Winner selection and prize money processing will be handled by General Chair of GEFCom2014 (Dr. Tao Hong, hongtao01@gmail.com).

NOTE: The contestants are NOT allowed to use external data other than what's provided in this competition. One exception is the load forecasting track, where the contestants can use U.S. federal holidays (http://archive.opm.gov/Operating_Status_Schedules/fedhol/2014.asp) in the models. Violation of this rule will result in disqualification.

Criteria

For each time period throughout the forecast horizon, the participants should provide the 0.01, 0.02, ..., 0.99 quantiles -- call 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 are 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:

$$L(q_a, y) = (1 - a/100) * (q_a - y), \text{ if } y < q_a; \\ a/100 * (y - q_a), \text{ if } y \geq q_a$$

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

To evaluate the full predictive densities, this score is then averaged over all target quantiles, from 0.01 to 0.99, for all time periods throughout the forecast horizon for all zones. The lower the score, the better the forecasts are. (see Rob Hyndman's blog for a more equation-friendly version: <http://robjhyndman.com/hyndsight/gefcom2014/>)

Every week, the entry with the lowest score is automatically selected as the weekly score for the team. The final score is the trimmed mean of the weekly scores. The highest and lowest weekly scores are discarded. The participants are required to submit entries for at least 9 out of 12 weeks during the scoring period to be eligible for a position in the final leaderboard.

Among the teams that beat the benchmark, up to 3 teams from each track will be recognized as winning teams to receive the team prizes