Data Description

In each of the 5 data files, there is a header row. Three columns of calendar variables: year, month of the year and day of the month. The last 24 columns are the 24 hours of the day.

In "Load\_history.csv", Column A is zone\_id ranging from 1 to 20.

In "Temperature\_history.csv", Column A is station\_id ranging from 1 to 11.

In "submission\_template.csv", "weights.csv", and "Benchmark.csv", Column A is id, the identifier for each row; Column B is zone\_id ranging from 1 to 21, where the 21st "zone" represents system level, which is the sum of the other 20 zones.

"Benchmark.csv" shows the results from a benchmark model.

Please make sure the submission strictly follow the format as indicated in "submission\_template.csv", where the year was sorted in smallest to largest order first, then month, then day, and then zone\_id.

This is the Load Forecasting track of Global Energy Forecasting Competition 2012 ([GEFCom2012](http://www.gefcom.org/)).This competition will bring together the state-of-the-art techniques for energy forecasting, serve as the bridge to connect academic research and industry practice, promote analytics in power engineering education, and prepare the industry to overcome forecasting challenges in the smart grid world.

The prize pool for the load forecasting track is $7,500. GEFCom is not a paper contest. Instead, this is a competition that requires participants to develop models and submit forecasts based on a given data set. Accuracy of the forecasts will be one evaluation criteria. In addition to accuracy, the participants are also required to submit a report describing the methodology, findings and models. Selected entries will be invited to IEEE PES General Meeting 2013 at Vancouver, Canada to present their methodologies and results. The team that finishes top of the leaderboard will win a cash prize. However an overall winner of the competition will be determined by the GEFCom Award Committee after the presentations based on forecasting accuracy, clarity of documentation, rigors of the approach, interpretability of the models and practicality to the industry. A few winning entries will be invited to submit the report in scientific paper format to prestigious scholarly journals, such as International Journal of Forecasting and IEEE Transactions on Smart Grid.

The topic for the load forecasting track is a hierarchical load forecasting problem: backcasting and forecasting hourly loads (in kW) for a US utility with 20 zones. The participants are required to backcast and forecast at both zonal level (20 series) and system (sum of the 20 zonal level series) level, totally 21 series.

Data (loads of 20 zones and temperature of 11 stations) history ranges from the 1st hour of 2004/1/1 to the 6th hour of 2008/6/30.

Given actual temperature history, the 8 weeks below in the load history are set to be missing and are required to be backcasted. It's OK to use the entire history to backcast these 8 weeks.

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;

2006/11/22 - 2006/11/28;

In addition, the particpants need to forecast hourly loads from 2008/7/1 to 2008/7/7. No actual temperatures are given for this week.