

NEM Data Challenge

2nd phase guidelines

DATA PROVIDED BY NEM SOLUTIONS

NEM Solutions will provide the candidates with two data sets:

- Historical data set: to be used by the candidates to define, train and test their proposals.
- Evaluation data set: to be used by NEM Solutions to evaluate the goodness of the candidates' proposals.

Historical data set description

The historical data set is formed by two types of files: files with raw information from the assets and files with weather forecasting. Both types of files will contain information corresponding to one month. The historical data set will provide 5 months of data (April - August 2015), and hence, five raw information files and 5 weather forecasting files will be provided in separate folders.

Raw information from the assets

The raw information data will be provided as CSV files of approximately 10GB each in “vertical” format (asset, date, time, variable, value) with a secundal data frequency, and will contain the raw data extracted from five wind turbines belonging to the same wind farm. Each file will contain information from the five assets for a period of one month.

As in the 1st phase, an XML file will be provided with the variables mapping.

The historical data set is similar to the one the candidates were asked to process in the 1st phase. Candidates will likely use the solution proposed during the 1st phase to “horizontalize” the dataset calculating ten-minute averages. Since the volume of data in the 2nd phase is larger, those candidates having troubles to process the data could contact NEM to ask for an already daily processed data set. This additional dataset will not be available until one week after the 2nd phase starting day.

Weather forecasting

The weather forecasting data will be provided as CSV files in “horizontal” format (asset, timestamp, range, wind speed, ambient temperature) with a hourly data frequency. The timestamp column indicates the time at which the weather forecasting was obtained in UNIX time. The range column indicates the forecasting time as the number of seconds from the timestamp.

Evaluation data set description

The evaluation data set is formed by several sets of two files corresponding to different evaluation journeys:

- A CSV file containing 10-minute averages of the raw information in “horizontal” format, extracted from the five assets during a single journey (24 hours).
- A CSV file containing hourly weather forecasting for the next 7 days to the journey.

The candidates will be asked to make use of the evaluation data set to generate the deliverables for their submission to NEM.

DELIVERABLES

The candidates should develop a module capable of estimating the energy a wind turbine will produce during the next 7 days. The candidates should deliver the following items before the deadline:

- Energy predictions: for each journey in the evaluation data set, candidates should provide a CSV file in “vertical” format (asset, range, energy production), with the **10-minute predicted energy production of the five assets during the next 7 days**.
- Source code: the source code implementing the module.
- Brief report: a PDF file, no longer than 5 pages, describing the proposed approach.

Note: if the candidates follow several approaches to predict the 10-minute energy production, they can send the “Energy predictions” for each approach in separate files, and describe the approaches in the “Brief report”.

EVALUATION CRITERIA

The candidates will be evaluated according to the following two criteria:

1. Prediction accuracy: several evaluation metrics such as the sum of squared residuals, the coefficient of determination (R^2), etc. will be calculated on the Energy predictions. To that effect, NEM disposes of the actual Energy production measurements corresponding to the Evaluation data set. The actual Energy production measurements will not by any means be accessible to the candidates until the second phase is over.
2. Clarity of exposition: the capacity of the candidate to clearly explain the proposed approach in the Brief report.