PhD Course- CONTROL AND OPTIMIZATION IN SMART-GRIDS

Session No. 3 –Renewable sources estimation - Exercise

In this session we are going to analyze real renewable sources data and to explore the behavior of this sources in a competitive electricity market. The aim is to evaluate the performance of a simple market, where a wind farm is price taker and bids firm energy.

First, data sets with prices, wind speed and solar radiation measurements are provided. From this data you should comprehend how these signals behave in a stochastic sense and gain information form the provided samples.

In the second part, you will obtain optimal bidding strategies for a wind plant subject to the wind speed time series provided before and evaluate its performance when participating in a competitive market.

Procedure

Part 1. Data analysis

Price Data

Market energy prices from one of the east-cost power systems in the USA has been obtained from the system operator PJM:

http://www.pjm.com/markets-and-operations/energy.aspx

1. Load price data from "*prices.mat*". This fil contains hourly registered prices for the day-ahead and real-time markets in the PJM region for the month of March 2018.
2. Plot the daily behavior of prices for the available 31 days. On average what is the difference between DA and RT prices?
3. Are peak prices occurring at the same hours in both markets?
4. Generate an average price curve for the month and obtain basic descriptive statistics. Which price presents more variability (uncertainty)?

Wind Data

The data set has been obtained from NREL National Wind Technology Center (NWTC) in Boulder, Colorado, USA. Minute-sampled data is available since 1996, for solar irradiation, wind speed and direction, temperature, among others. You can access the database at http://midcdmz.nrel.gov/nwtc_m2/ .

1. Load data *set wind_data.xlsx,* this is a spreadsheet with hourly averaged wind speeds at 50 and 80 meters above ground.
2. Plot the first 10.000 samples of wind speed time series.

   a. What is the mean speed?
   b. Do you identify outliers?
   c. How can we clean the set?
3. Eliminate extremely low values and build a histogram of the series, does it behave as a Gaussian distribution?
4. Construct an empirical probability density function (ecdf). Can you parametrize it?
5. Fit a Weibull distribution to the data, use de available Matlab function (wblfit). Does it properly represent the available data (*ecdfhist, wblpdf*)?

Solar data

The data set has been provided by Bogota District Environment office. Hourly registered solar radiation samples for the complete 2015 year are available for 5 points within the city. Outliers and missing data have been already removed.

1. Load data set from file "*solar_data.mat"*
2. Obtain the correlation function between the 5 time-series, up to 50 hours of lag. Do you identify any two points that show higher correlation?
3. Select one of the time series and create two System Identification Toolbox data structures (*iddata*). One using the first 70% of data and a second on with the remaining 30%.
4. Open the System Identification GUI (*systemIdentification*) and load the structures created before.
5. Remove the mean value of the signal and estimate multiple polynomial models with different regressor lengths form the data set with the first 70% of data.
6. Obtain the one step ahead prediction and FIT on the same dataset and also on the 30% of data saved for validation.
7. Repeat the previous analysis for 5, 10 and 20 hours ahead prediction horizon. How does the estimator behave as the horizon is increased?
8. Obtain the residues (estimation error) autocorrelation function. If the model properly explain data, the error should be white noise. Does the resulting error sequence have an autocorrelation function corresponding to a white noise signal? Can you improve the performance by varying the regressor size?
9. Export the best ARX model you can find and leave the System Identification GUI.
10. On the command window, build identification (70%) and validation (30%) sets (*iddata*).
11. Estimate non-linear AR models (*nlarx)* for one-hour ahead prediction, use a sigmoid NN as model structure. Vary the regressor length considering the same orders of magnitude used for the linear case.
12. Select a model with adequate regressor length and evaluate the predictor performance for a 12 hours-ahead prediction on validation data. Does it provide better performance that a linear model?

Part 2. Market participation

Assume that you are the manager of a wind farm located at the region where the wind speed dataset has been collected. Based on the stochastic model of the wind time series and the price information used in the first part decide how much power to bit in the day-ahead market and evaluate the performance of the strategy.

1. In general, the power output of a wind turbine increases as the third power of wind speed. Obtain the time series of the third power of the wind speed and plot a histogram, does it exhibit the same shape of the original samples? For simplicity assume that output power can be explained with the Weibull distribution estimated in the first part, normalizing the signal to 1 as the rated peak power.
2. Given the results of the prices analysis in Part 1, build the regions on the plane of imbalance prices that define the optimal strategies of capacity bids for each hour of the day.
3. Assume that there are not penalties for positive imbalances (extra generation is withdrawn). For each hour of the month evaluate the optimal bidding capacity, based on the known energy price and negative imbalance expected value *per hour*.
4. Generate random samples from the Weibull distribution estimated before and, for each hour, calculate the income, penalty, profit and deviation for the generator. Is it profitable to participate in the market?
5. How much reserve power should be scheduled, on average, by the SO to compensate the deviations caused by the random wind generator? What is the worst-case deviation? How much power is withdrawn during the month?