Wind Analysis Homework

Elliot Cohen, Sustainable Engineering Lab

September 16, 2014

Problem Statement

Suppose you are hired as a consultant to evaluate the potential for wind energy production on a small chain of islands in the Atlantic Ocean. The islands already have one wind farm up and running. Pertinent project specs are as follows: - **Power curve**: Relationship between wind speed and power output for a wind turbine. - 7 Vestas wind turbines (v52-850Kw) rated @ 850kW each. - Combined rated capacity of the 7 turbines is 5950 kW. + This is roughly 100% of the annual average power demand for the island. + This is a huge penetration of wind! - 10-minute average wind speed and cummulative wind energy production data are provided by the utility for the most recent year, in .xlsx spreadsheet format. - Hourly load was estimated in-house from the load profile of a nearby island, and is provided in a separate .xlsx spreasheet.

Using this data, please prepare the following for an investor meeting coming up next week! Be sure to document any assumptions, equations or algorithims implimented, and include **all** underlying R code in a well-commented Rmarkdown file, [knit(http://yihui.name/knitr/) (e.g. rendered) into a clean .html or .pdf file.

1. Visually inspect the data!
    - Does it make sense?
    - Are there any obvious issues, such as missing data, incomplete records or suspect values?
2. Summarize the data in a few simple ways to make sure the values make sense (use benchmark comparisons!).
3. Clean the data. There are two schools of thought on this:
    - Option 1: Remove all records containing suspect data (e.g. if the wind and/or power data at a given timestamp are not possible on their own *or* do not make sense together as a pairwise observation. Could you produce 400kw at 0 windspeed?).
        - Use textbook knowledge of wind energy systems as a simple check (hint: Betz Limit!)
    - Option 2: Data correction, again using textbook knowledge of the system.
    - After implementing step 3 (option 1 or 2), you should have a "clean" dataset. Now make engineering computations with confidence! Proceed with steps 4-9 (and 10-11 if you're extra ambitious!)
4. Total wind energy produced last year (e.g. KWh delivered to the grid).
5. Capacity factor of the current system.
6. Total wind energy that was *possible* given the observed windspeeds and the turbine power curve. This is the uncurtailed power output.
7. Uncurtailed capacity factor using the result from 6 (this assumes the grid could accept all wind power output).
8. Turbine efficiency:
    - average for the entire windfarm; and
    - as a function of windspeed.
    - compare with the Betz Limit.
9. Characterization of the wind resource using a Weibull distribution.

10. **BONUS** Randomly generate a year's worth of new windspeed data according to the fitted Weibull distrubution. Using the new windspeeds, predict what the resulting wind energy production may look like next year.

11. **BONUS**: Repeat step 10 500 times and compute the capacity factor each time. Boxplot the results of the 500 simulations. This is called an ensemble forecast.

Now suppose you have reason to beleive that the anemometor (wind speed gage) was off by a few meters per second, for a few months of the year. The system engineer tells you that the anemometer was reading high. What does that mean? How could you correct the windspeed? (hint: use the manufacturer's power curve). Did you observe this issue in step 3? How did you handle it? If in step 3 you chose to remove all suspect data, that's fine, but now go back and try a windspeed correction. If you already did a windspeed correction, go back and try removing the data. Compare the results.

Everyone should now have completed steps 3-9, twice! Once with each data cleaning option.

Additional instructions

This bulk of this assignment (steps 1:9) can be done in Excel. This is often a good starting point, as it allows you to focus on the engineering concepts, not programming concepts. Steps 9-11 are far easier to do in R or another scripting language (e.g. Python, Matlab, etc…) than it would be to do in Excel, if at all possible.

You may do the assignment in either platform. We recommend staring in Excel to see how far that takes you, and then move into R for more advanced (and faster) analysis and vizualization.

Whatever you do, be sure to COMMENT your work clearly and succinctly.

Key concepts covered in this assignment

- **Power curve**: Relationship between wind speed and power output for a wind turbine.
- **Weibull distribution**: Characterization of the wind resource for a given location (e.g. estimate parameters for a statistical distribution fit to the wind speed data)
- **Capacity factor** (optimal and actual): Energy output / (nameplate capacity x time)

Hints if you get stuck:

Useful functions in Microsoft Excel

- **VLOOKUP**(lookup value, array, col num, [lookup range])
  - Lookup a value in a column of an array based on the closest matching value in another column of the array.
  - e.g., each row of the Manufacturer's Power Curve array contains a windspeed and corresponding power output. Use VLOOKUP to supply a random windspeed, query the array and return the corresponding power output.
- **IF** (condition, statement, else if, statement)
  - Useful conditional with many applications
- **FREQUENCY** (data, bins)
  - Counts the number of observations that fall in each bin.
- **WEIBULL.DIST** (value, shape factor, scale factor)
  Estimates the probability of a value given the fitted parameters of a Weibull distribution.

Useful functions in R:

- **str**(df)
  - Look at the structure of any object
- **head(df)**
  - Look at the first 6 rows of an array-like object (especially useful for data.frames and arrays).
- **read.csv**("file_name", optional_arguments)
  - For reading .csv data into R.
- **subset**(df, column_name [operator] criteria)
  - Remove records (e.g. rows) that do not meet some criteria.
- **na.omit**(df)
  - Omit records (e.g. rows) with missing values.
- **which**(column_name [operator] criteria)
  - Returns the row index of records that meet some criteria.
- **plyr::ddply**(df, .(variables), function)
  - For each subset of a data frame, apply some function then combine results into a data frame.
- **ggplot2::ggplot**(df, aes(x, y, , other aesthetics))
  - ggplot() is typically used to construct a plot incrementally, using the + operator to add layers to the existing ggplot object.

Data extraction from .csv

Extract only the data you need from the raw data: - Total Energy Delivered to the grid in ten minutes (measured); - Total Active Power averaged over ten minutes (measured); - Total Possible Energy output in ten minutes given measured windspeed (calculated by SCADA); - Total Possible Energy output in ten minutes given measured windspeed (calculated from the turbine Manufacturer's Power Curve).

Remove or re-assign (obviously) erroneus data

Correct the Energy Delivered measurements. We re-assign all negative values to zero, and re-assign values greater than what is possible given the rated capacity of the turbines to the rated capacity.

Alternatively, you could subset the data, removing all records associated with suspect values. However, this can unessesarily shrink the amount of data you have to work with. As a general rule of thumb, strict filtering (e.g. subsetting) is preferable when you still retain ample data for whatever modeling or analysis is to follow. By contrast, when you are data limited, think twice before applying stringent filters.

In either case, you are introducing some bias to the data. It is important to test the sensitivity of your results to data filterning/correction. Try it both ways (filtering/no correction; correction/no filtering) and see how the results differ.

Manufacturer's power curve

The manufacturer's power curve [MPC] specifies the design output (kW) for a specific model turbine at a range of windspeeds in the operating range of the turbine. The curve is discretized, so you must use a lookup function with a certain tolerance, *or* fit a smoothing spline (or polynomial) to the data to make it continuous. Once you have this curve, you can supply any windspeed within the operating range of the turbine, and estimate the power output.

Windspeed correction

Correct windspeed measurements. If we assume the Energy_Delivered is correct, then we can lookup the corresponding windspeed required to produce that much energy in ten minutes according to the Turbine Manufacturer's Power Curve. If the measured windspeed is less than the windspeed according to the Power Curve, then we re-assign it be the windspeed value from the Power Curve.

As an example, we know that it is impossible to produce 400kw at 1mps windspeed. [In fact, the turbines switchoff at windspeeds below 3mps and above 25mps.] So, if we see a value of 400kw produced at 1mps, we can assume the wind measurement is inaccurate. [This assumes the energy measurement is correct, which is a decent assumption consdidering there are multiple points at which energy can be measured throughout the system, e.g. ex-bus bar, load dispatch center, etc.]

To correct the windspeed, we look up the windspeed cooresponding to 400kw on the manufacturuers power curve. Here we see the windspeed must have been at least 8.5mps. We re-assign the suspect windspeed to that value.

Computing capacity factor

Compute the acutal (curtailed) capacity factor, as well as the possible (uncurtailed) capacity factor according to the MPC and SCADA Possible_Power measurements, respectively. Recall that capacity factor is simply the total amount of energy generated in one year divided by the installed capacity of the system times the number of hours in one year.

`CF_actual = Energy Delivered / (nameplate capacity x time)  CF_optimal= Energy Possible / (nameplate capacity x time)`