Report on Statistical Analysis of Capstone Data

The Haute Borne Wind Farm

Max Davis

Some research into energy production revealed that a good measure of turbine performance is Active Power. This measure can generally be considered equivalent to how much power the wind turbine is producing. Since the dataset contains an averaged Active Power for increments of time throughout each day in period for which there is data, it seemed like a good place to start exploring.

As this is data from Wind Turbine Generators, I started with the basic question: “What is the relationship between wind speed and active power?” To begin to answer the question, I made a plot with wind speed on the x-axis and Active Power on the y-axis, which did indeed show a positive relationship, albeit with some interesting anomalies. To better understand the anomalies, I facetted the plot by individual wind turbine. Since there are four turbines in the data, this meant four facets, each with its own plot comparing Active Power to Wind Speed. Two of the turbines had a more pronounced deviation from the main trend of the data. This was in the form of a steeply positive sloping branch to the left of the main branch.

To keep the same basic visualization and try to learn what made those areas different, I started mapping different variables to color, and found that by mapping date to color, it was apparent that the break away groupings in each of the turbines mostly seem to be in the chronologically earlier part of the sample, appearing as dark blue spikes to the left of the main grouping (where dark blue corresponds to earlier dates). Could it be the turbines ran more efficiently sooner after installation? However, the very highest points of efficiency (where there is wind speed around 5 m/s, and high active power) are later in the time series, or lighter in color. This plot also raised another question: why might turbine R80711 have a significantly larger range of active power readings at the higher wind speeds? Above ~12.5 m/s, the correlation between wind speed and active power becomes much weaker for that turbine.

To explore the question having to do with a spike in efficiency earlier in the data, I tried limiting the dates to the beginning of 2017, since that was the time period where the outliers are clustered most distinctly. For the two dramatically deviating wind turbines (R80790 and R80711), the seeming spike in active power is apparently roughly when the wind is between 3 and 6 m/s. More investigation would be necessary to determine how significant these clusters of data may be. Also apparent from this new plot is that in three out of four of the turbines during this period, there are many data points at which zero Active Power is recorded, and at wind speeds when we would expect to see Active Power increasing.

To get a different perspective, I tried comparing wind speeds over time, and how the active power is affected. For a new set of plots, I mapped the date on the X-axis, and wind speed on the Y-axis, and this time mapped Active Power to color. Over a full year period, the resulting plot showed that wind rise and fall over time is cyclical, with a generally corresponding increase and decrease of Active Power as observed in the previous plots. But in this plot, some dark spikes indicate low to zero active power at high wind speeds.

As an experiment, I isolated the darkest spike in Turbine R80790’s time series. Since this dark spike represents an unexpected lack of Active Power at high wind speeds for a given amount of time, I filtered out a smaller date range on either side of the dark spike (9 days) and changed the date label to 1 day instead of 6 months to see exactly which days are in question. The resulting plot revealed that the dark spike occured over just a two day period, February 3rd and 4th of 2017, when despite the windy conditions, and the three other turbines producing power, Turbine R80790 was not producing. Further exploration might determine whether it was down for service (would require access to service records), or perhaps we can read into the data to figure out if something else created a fault and if there might have been a predictor for that fault in the preceding time periods.

Next steps will be doing some more exploration, and hopefully identifying what might be some predictors of a turbine recording high wind speeds, but not recording any Active Power. If these predictors could be identified in the data, we might be able to offer a client suggestions for anticipating faults and proactively responding to them.