Anomaly Detection in a Wind Farm using Double Measurements

José-Antonio Rodríguez-Gallego, Department of Differential Equations and Numerical Analysis, University of Seville

As part of the research made under the ARIA (*Accurate Roms for Industrial Applications*) project there is a primary need for reliable data in order to validate the different Reduced Order Models (ROMs) within the project. However, the results with the available data can be controversial as substantial differences appear when using different power production measurement devices. Therefore, the need for data analysis techniques is crucial to filter out these differences and derive reliable data. Furthermore, the use and development of ROMs is essential as a tool for determining the dependency of power production on different parameters, such as wind speed, blade pitch or wind variation over time.

Considering the increasing number of wind farms appearing worldwide and their size it is mandatory to take into account the expected power production that comes with them. To do so, one of most accessible tools available is the analysis of the Contractual Power Curve (CPC) and the Experimental Power Curve (EPC), which in both cases are graphs comparing mean power production at different wind speeds.

One of the limits this analysis shows, as with many other problems, is the dependance on the available data, *i.e.*, the measured wind speed is given by the instruments built into the turbine, out of Wind Farm Owner's control. To increase control over the registered production and the values of the surrounding variables one option is adding another measuring system, therefore providing a second data source to compare the behavior of the different turbines and the involved variables.

As said, in this case we did not employ only the data given by the manufacturer, but also the data we have gathered by other means, therefore using in this case data from both the SCADA (Supervisory Control And Data Acquisition system) and the independent wind measurement systems.

After the analysis of the available data we decided to work mainly with one particular wind farm, "wind farm A", choosing the latter as our main focus for the analysis over other possible wind farms. The data originally was stored in different formats and came from different sources: the Energy Management System (EMS), SCADA and independent systems; therefore, it was necessary to clean and prepare the data by mixing it before performing any possible analysis.

After those steps and merging all the data into a single dataset of over 200.000 entries of 20 variables, we were able to perform an exploratory analysis of the data (EDA), focusing on turbines T3 and T7 due to their reliability, and emphasizing key variables and their correlations over time. After checking key behaviors, like the Weibull distribution of wind speed or the expected correlations between wind speed related variables and produced energy we could address the real objective, comparing the different registered wind speeds —after correction— and the produced power, both theoretically and in practice.

When executing the contrast of behavior between power production against wind for these two turbines we found the appearances of significant differences, with a mean absolute value of 0.59 and 0.57 m/s for T3 and T7, respectively, with a standard deviation across the samples of 0.25 and 0.34 m/s. Notice that these differences could range to up to 4.67 m/s in the most extreme cases.

Afterwards, by means of using a Random Forest model [1] and an ANOVA test over the subset for T3

and T7, we identified wind speed, blade pitch, wind speed variation over time, and yaw pressure as crucial variables influencing power production in general. The model demonstrated its suitability for this subject, as evidenced by low error metrics, with an R^2 exceeding 0.99.
After knowing of the existence of a possible anomaly in 2021 for one of the turbines, T1, we decided to analyze in depth the behavior of wind production in that period, comparing it with the already analyzed turbines. In this case we found differences in the measurements of wind speed between two measuring systems with a mean value of 0.45 m/s, which increased to a mean of 1.13 m/s after a certain moment, which we baptized *SC21*. By using an ANOVA test it was proven that this growth could have been related to power production and the blade pitch of the turbine during the period. Our hypotheses suggest, however, that one set of wind speed measurements was adjusted to align with the contractual power curve. To assess this, the following figure is included:

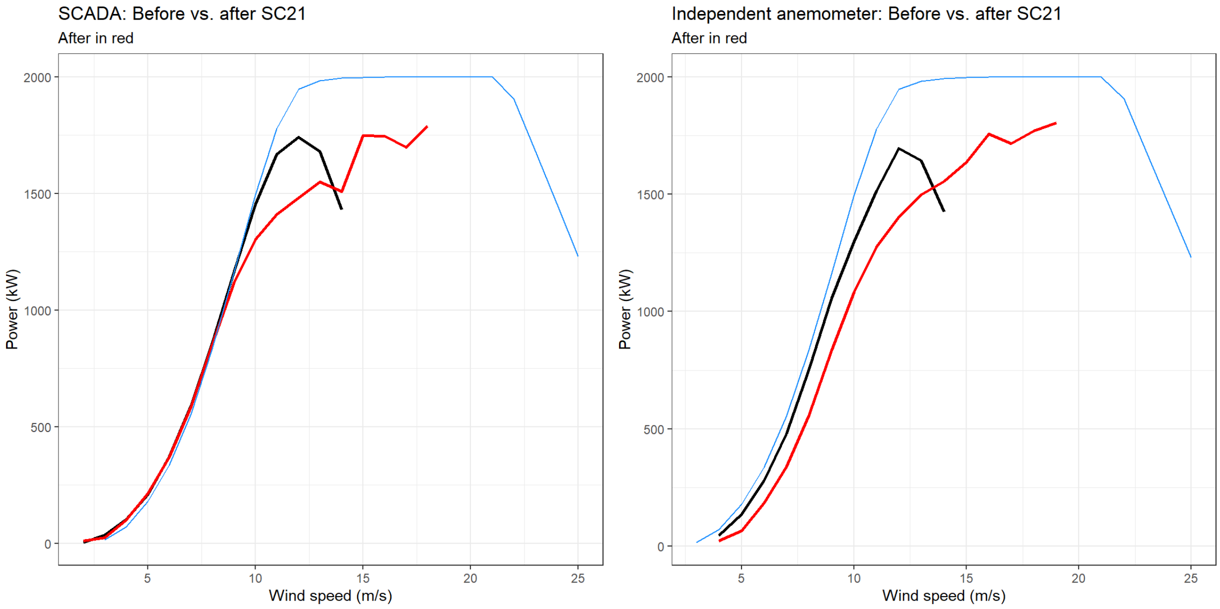

Figure 1: *Experimental Power Curve (EPC) from the independent anemometer and the SCADA systems, before and after* SC21*, the anomaly detected for turbine T1 at wind farm A. In blue we can see the Theoretical Power Curve (TPC)*.

As we can see, for the SCADA system the match between the EPC and the TPC is absolute, while for the independent measurement system the behavior is not completely equal to the TPC, and it even changes more after the *SC21* incident. Even if this shows how the different measurements could be related to the EPC, further analysis is still required.

With the previous results a complete analysis of the available data for wind farm A was carried out, showing out in particular the anomaly regarding power production at turbine T1 during 2021, as well as some of the variables that could have been influencing the change in power production.

As different behaviors were detected for turbine 1, future work will try to automatize this anomaly detection technique, expanding the result for a bigger myriad of possible cases. With this, the objective would be to detect these anomalies quickly, so that the causes for it can be easily detected and the losses minimized. The automated classification of anomalies in the general case is, however, much more demanding, as we don't have *a priori* knowledge about the expected "normal" behavior, apart from the TPC.

Bibliography

[1] Ho, T. K. (1995). Random decision forests. In Proceedings of 3rd international conference on document analysis and recognition (Vol. 1, pp. 278–282).