NEM DATA CHALLENGE, 1st Phase

Sergio Díaz Valdés, August 2017

Executive Summary: This report presents the analysis of the data concerning the condition monitoring measurements of five wind turbine generators (assets). The data contains 28 speed, velocity, temperature, direction, and pressure sensor signals sampled every second throughout May 2015, summing up to about 375 million individual values gathered in a 15Gb CSV-format file. As preparation for analysis, the data has been resampled and averaged every 10 minutes. After exploring the data by calculating descriptive statistics and by creating visualizations, several plots relevant to Wind Turbine Generators (WTG) have been produced.

Table 1. Summary statistics of selected variables for asset A001.

Variable Name	Mean	Std	Min	25%	50%	75%	Max
WNACWindSpeed	0.193	0.076	0.011	0.136	0.196	0.246	0.590
WNACAmbTemp	0.395	0.141	0.141	0.288	0.383	0.500	0.756
WNACNacelleTemp	0.598	0.124	0.314	0.507	0.592	0.682	0.892
WGENGenSpeed	0.751	0.207	-0.047	0.602	0.785	0.941	0.962
WROTSpeed	0.434	0.119	0.000	0.348	0.453	0.543	0.556
WTURReactivePower	0.475	0.012	0.456	0.466	0.472	0.480	0.515
WTURPower	0.336	0.286	-0.010	0.077	0.275	0.547	1.004
WHDRGroupOilPress	0.947	0.002	0.932	0.945	0.947	0.948	0.956
WCNVNetVoltage	63.03	0.48	61.45	62.69	63.03	63.37	64.76

Table 2. Missing values count in the recorded signals.

Asset	Date	From	To	Count
A001	26/05/2015	9:10	11:00	12
A002	21/05/2015	11:20	11:30	2
A003	25/05/2015	10:00	13:10	20
A003	27/05/2015	11:20	11:20	1
A004	18/05/2015	9:50	12:10	15
A004	21/05/2015	11:50	12:00	2
A005	06/05/2015	6:30	13:20	42
A005	07/05/2015	6:30	13:10	41
A005	25/05/2015	11:50	11:50	1

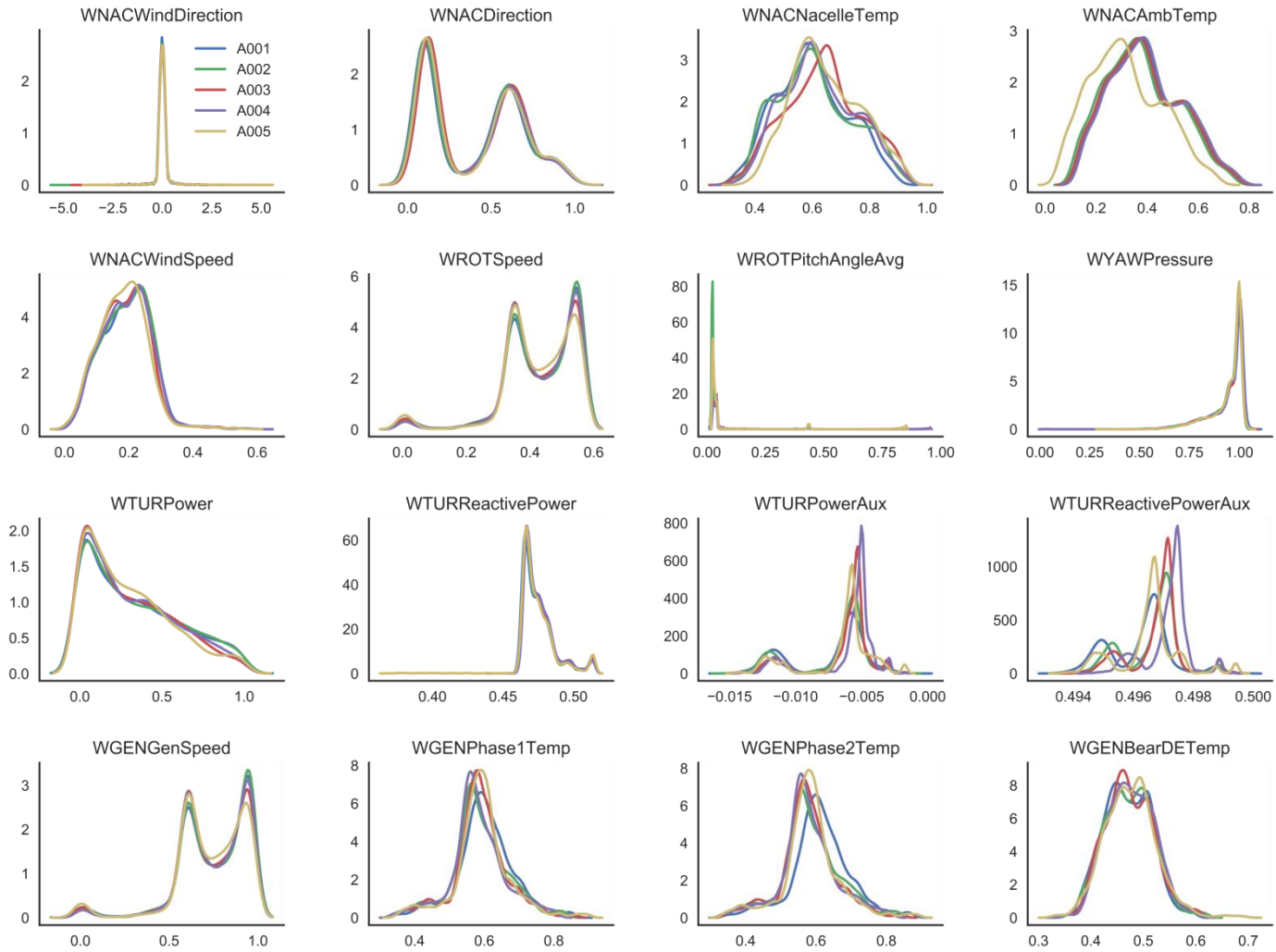

Figure 1. Kernel density estimation plots for selected WTG variables. For brevity, some sensor parameters were not inlcuded given their similarity to those shown above.

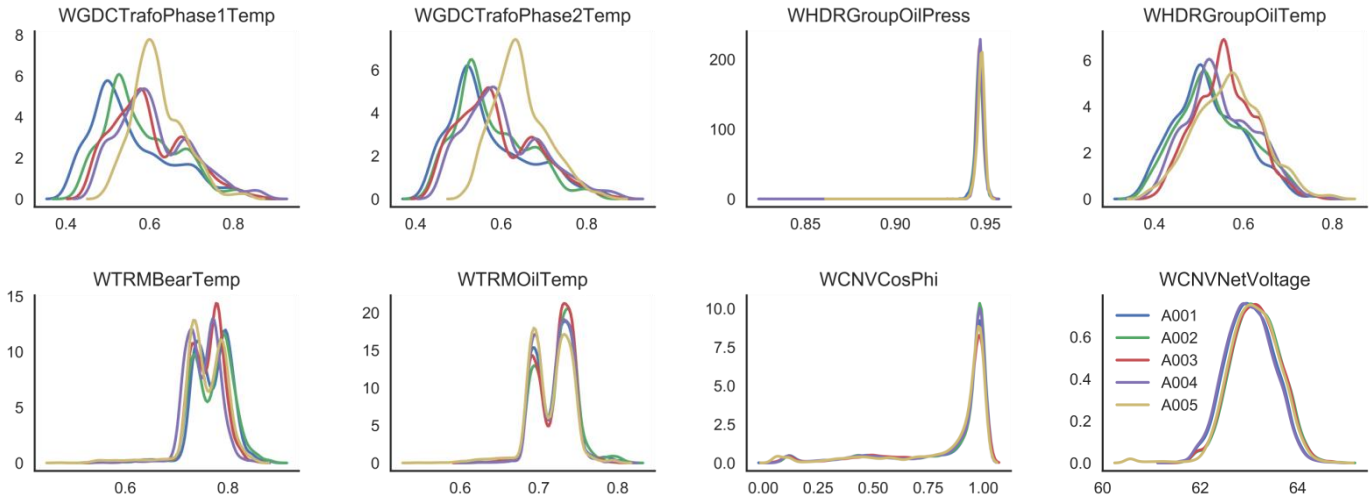

Figure 1 (continued). Kernel density estimation plots for selected WTG variables. For brevity, some sensor parameters were not inlcuded given their similarity to those shown above.

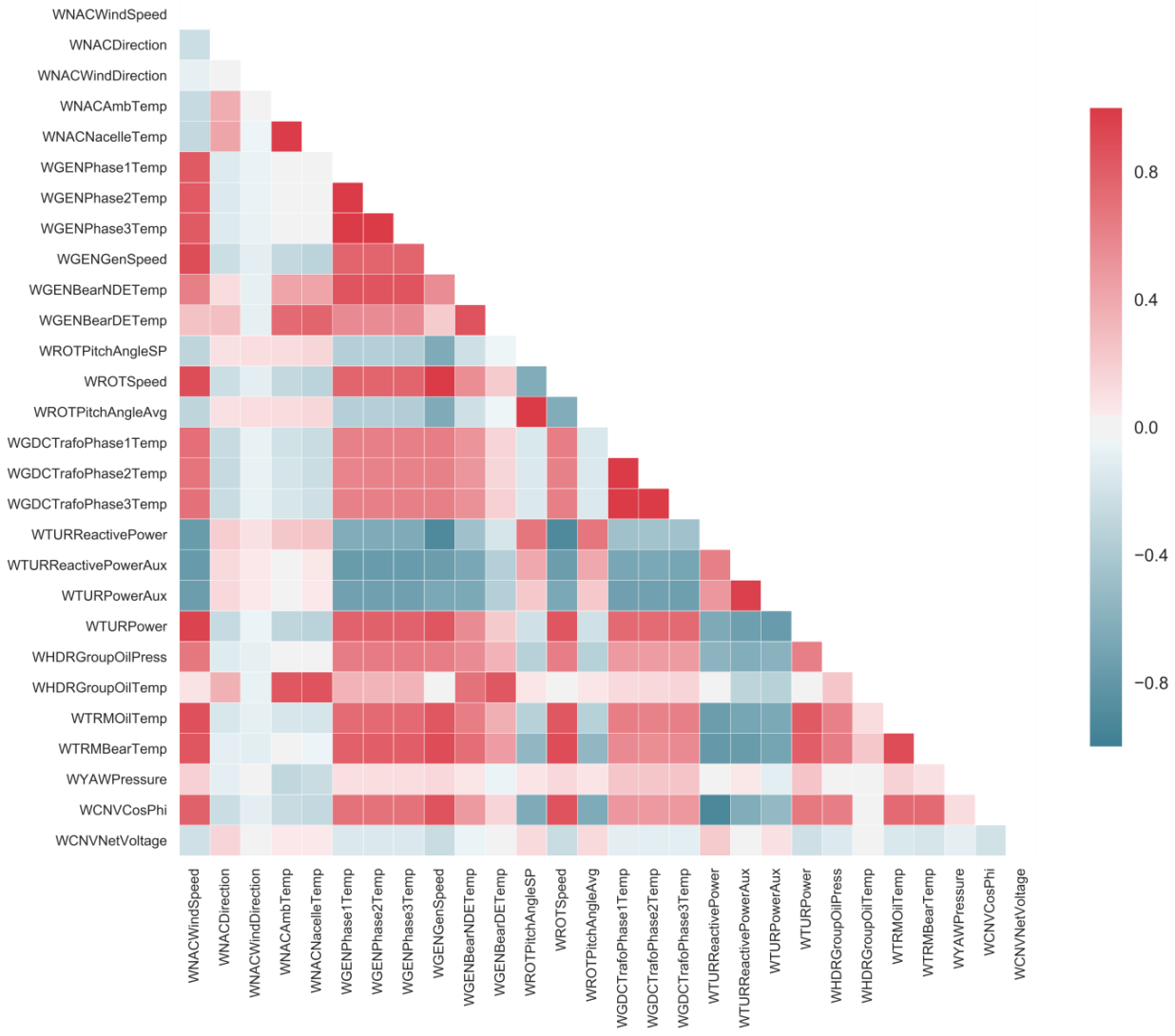

Figure 2. After exploring the individual variables, an attempt was made to identify relationships between features in the data using a correlation matrix, where values close to 1 or -1 are indicative of an apparent strong relationship between parameters.

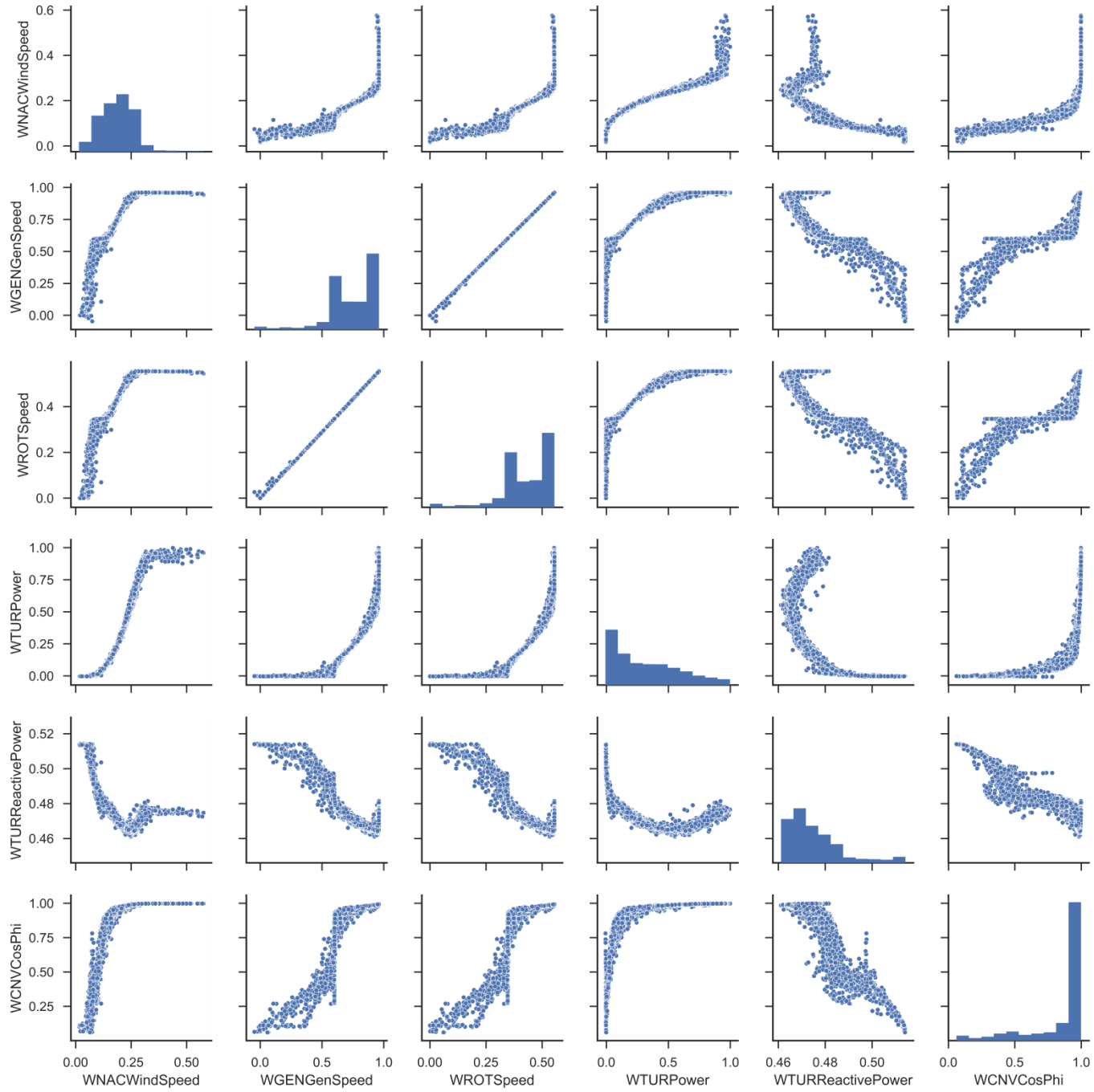

Figure 3. The figure is made of a number of pair-wise scatter plots for those variables that obtained highest scores in the correlation matrix. Due to space constraints, only six sensor variables are shown above.

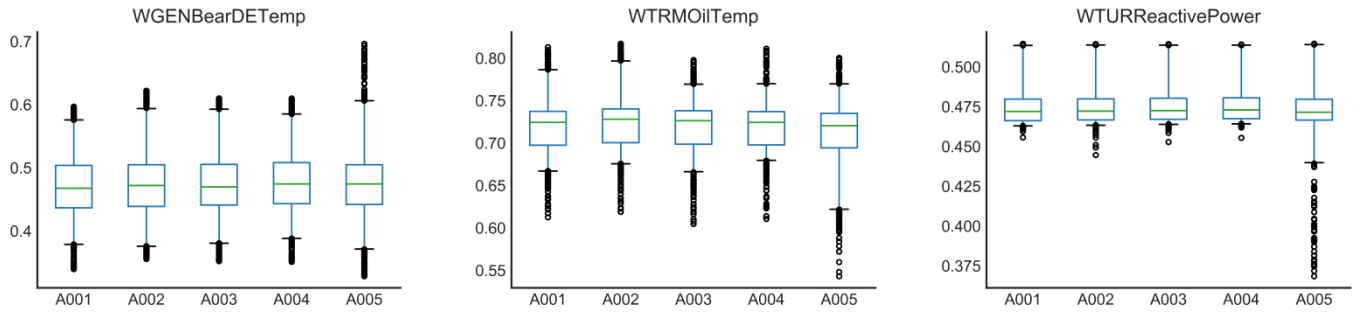

Figure 4. Box-and-whisker plots were generated for each sensor measurement to identify the presence of outliers. Plots where asset A005 shows higher number of outliers are shown in this figure.

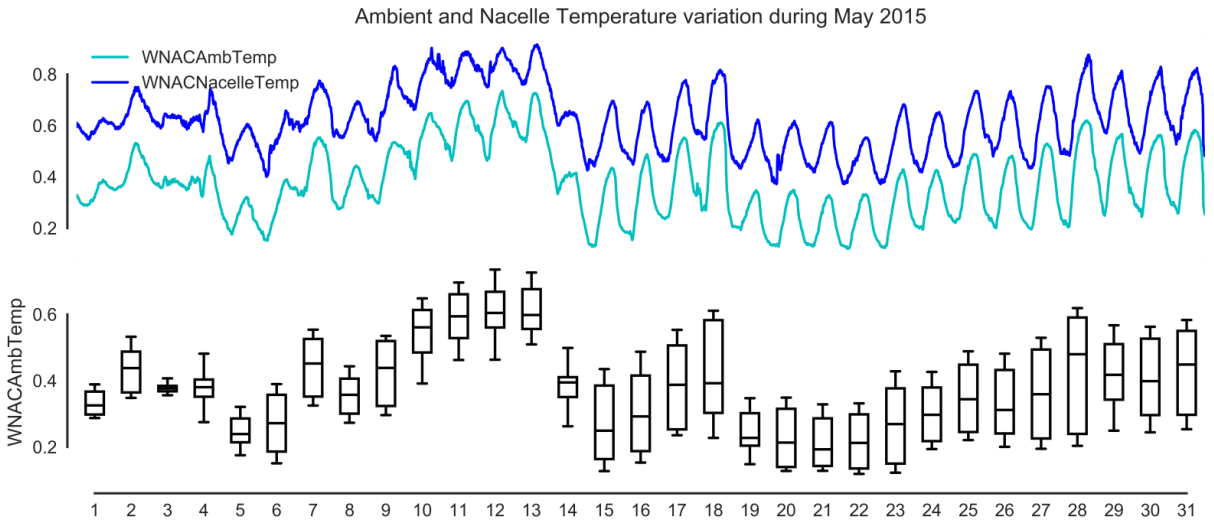

Figure 5. Ambient and Nacelle temperature variation throughout May 2015. Values shown represent the average across all assets.

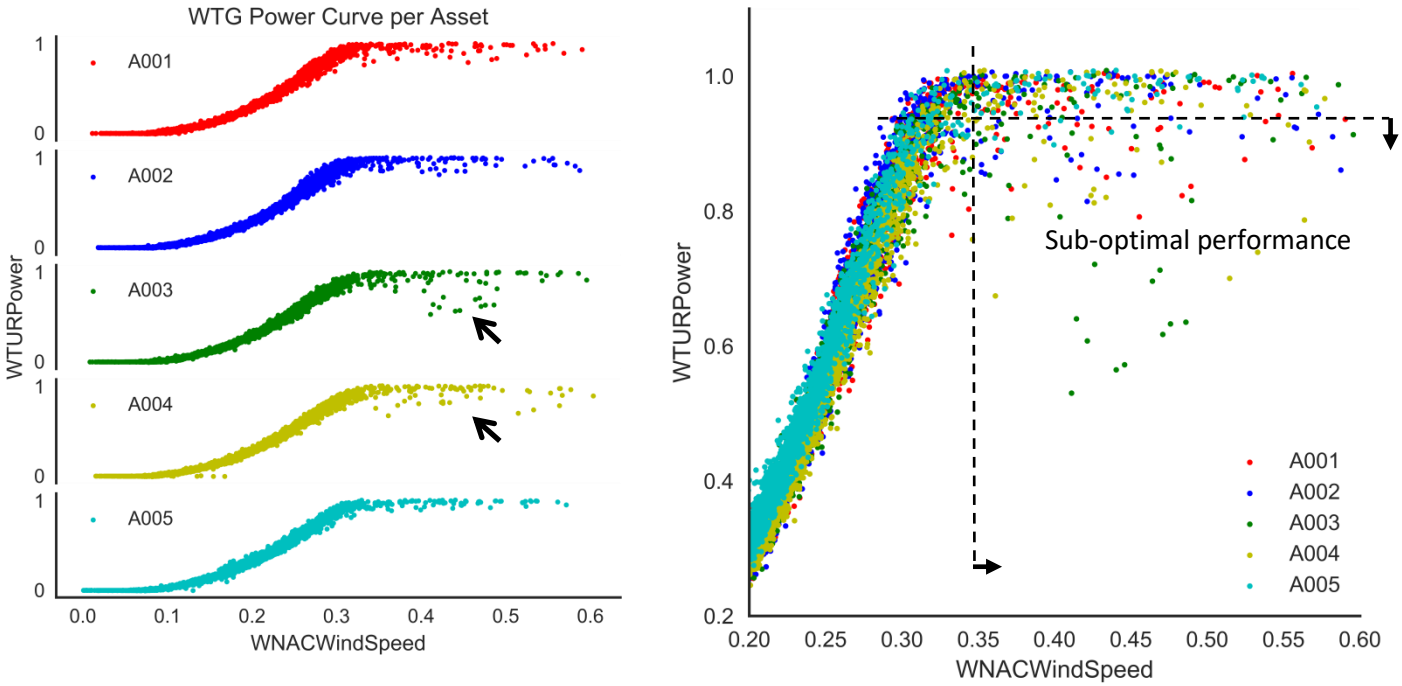

Figure 7. The figure on the left shows the complete characteristic Power Curve for all assets, where arrows indicate data points that require further inspection since they diverge from the expected trend. The image on the right reveals that all assets contain sub-optimal performance data points.

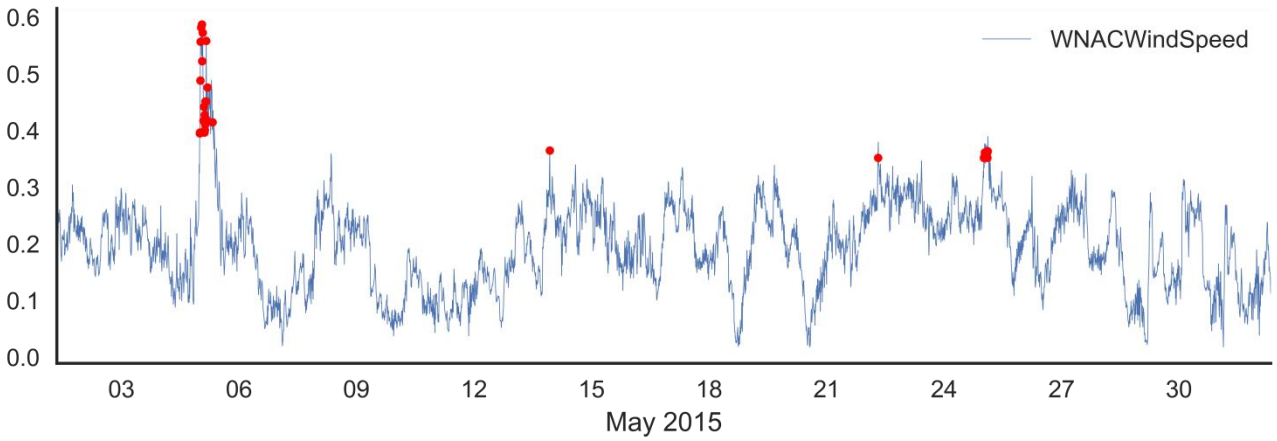

Figure 8. The measured wind speed signal for asset A002 is shown here, where red dots correspond to those data points identified in Figure 7 as sub-optimal, which in turn correspond to above normal wind speed values.

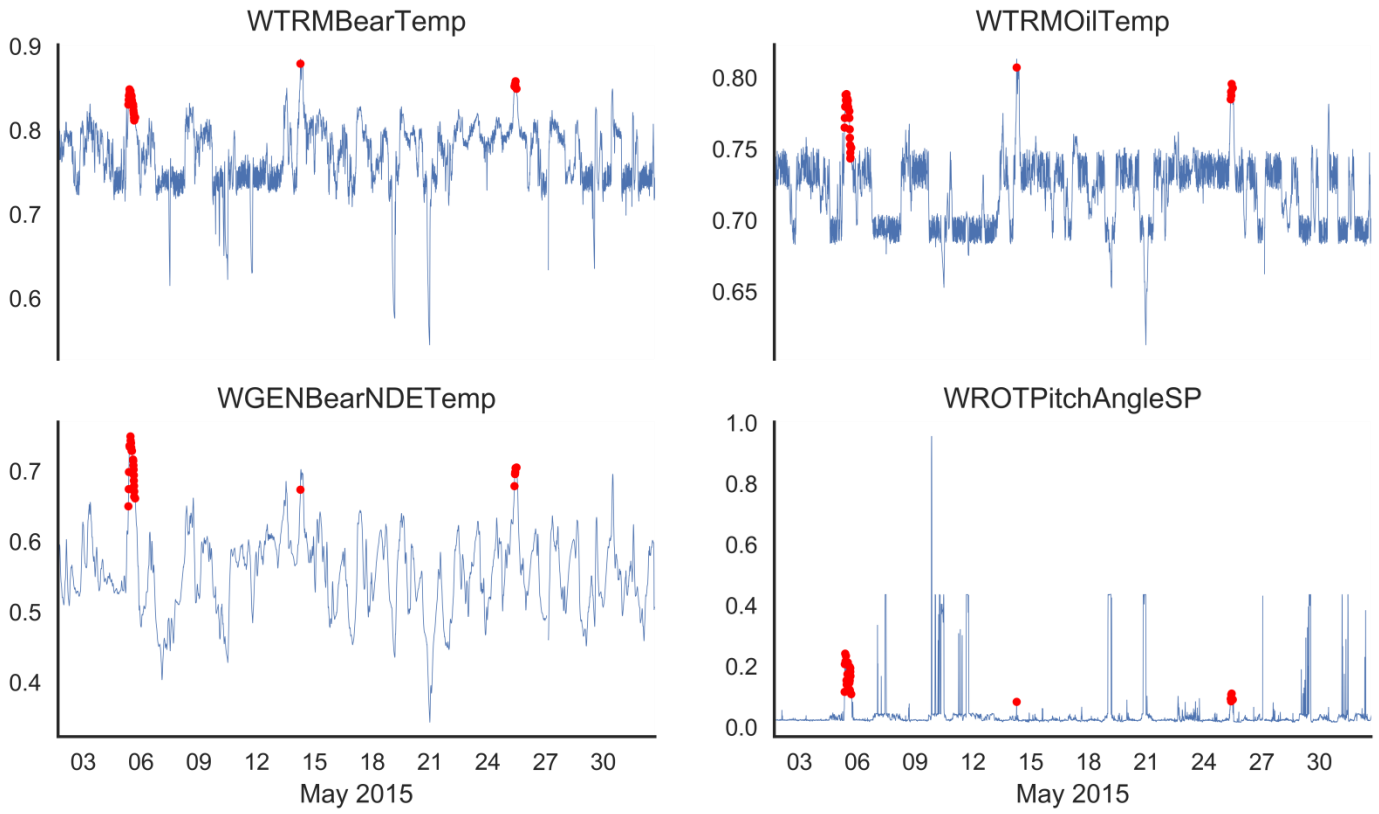

Figure 9. In connection to figures 7 and 8, asset A001 sensor measurement plots shown here also highlight the location in time of those low-performance data points. By inspecting the plots it becomes apparent that as the wind speed reaches a safety or cut-out wind speed or as a sensor measurmemte gets closer to an alarm-threshold value, the rotor is forced to reduce speed, probably using blade pitching, to avoid mechanical damage.

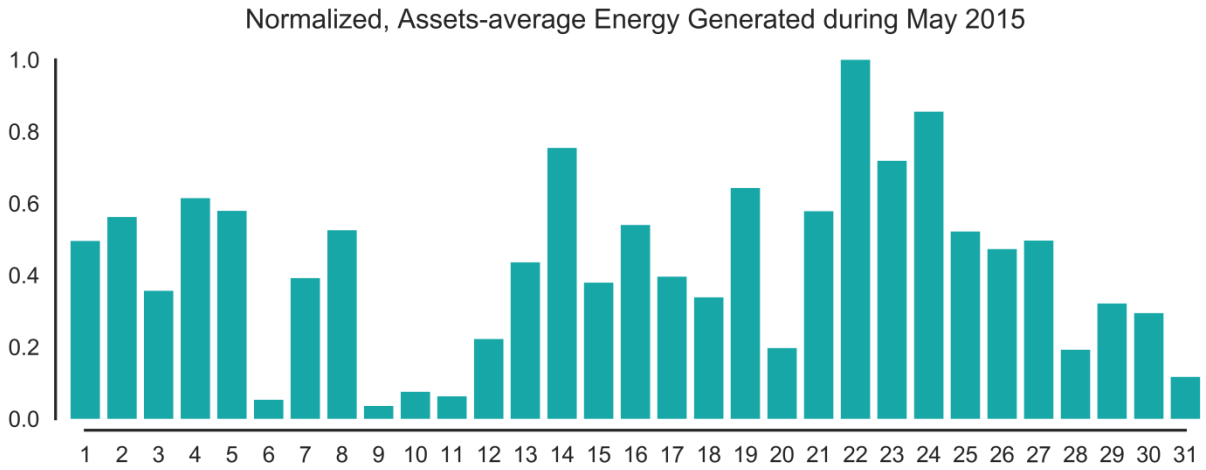

Figure 10. The daily power generated as shown here is the average of all assets and normalized with respect to the highest daily output in the month. All assets present similar plots with minor variations among them.

Concluding remarks

The WTG condition monitoring data has been successfully processed and analyzed using Microsoft Power Bi and a purposely developed Python language script. A selected number of visualizations have been produced to explore relationships among sensor measurements. The data contains valuable information on asset performance and can be used to develop health diagnostic and prognostic tools.