Finding a way to manage big data

Uncovering the connection between the environmental conditions at a wind farm and the effect on the turbines will enable better siting and more efficient maintenance regimes. Managing the data resource available is key, says Pieter Jan Jordaens and Jan Helsen of OWI-Lab

Not all wind sites are the same. Some sites wear out the turbines quicker than others, due to harsher loading conditions, greater wake effects or sporadically reoccurring special loading events. Think about it like two second-hand cars of the same make and vintage, with the same number of kilometres on the clock. But one has been driven gently with consideration for brake, suspension, tyre and transmission wear, while the other has been handled roughly, revved hard before the engine and gearbox oil have had time to warm up. To know its condition, you need to understand how a second-hand car has been driven, not just how far.

You can carry the car analogy further, to accidents. Insurers only have limited feedback about driver behaviour, being aware only of the incidents reported, not the near-misses or small scrapes not worth a claim. If they had access to data offering insight into the driver's style and habits, they might have a different assessment of the chances of that driver being involved in a crash.

"Big data" is increasingly setting the scene for this type of analysis in different industries. For wind power, its use is crucial in setting the industry on the pathway to reducing the levelised cost of energy (LCOE).

Detailed knowledge of a turbine's load history allows a replay of certain events to learn how the turbine responded to them. This enables improvements in drivetrain design that originate from feedback on the existing fleet and more detailed validation of advanced simulation models used in the design phase.

Future perspective

Data analysts interpret this information from two perspectives: historical and future. Based on historical data, key figures on performance, reliability, efficiency and special events are identified to help improve existing designs and optimise the layout of new wind farms. The future perspective investigates the data to try to predict what would happen, and it is this mindset that shows the most potential in optimising current operations and maintenance (O&M) regimes.

As an offshore wind research organisation, OWI-Lab has worked with Belgian wind project operators since 2011, looking for gains that can be achieved in O&M.

We combined the high-frequency sampled measurement data gathered from selected turbines with the traditional available turbine Scada-parameters and data on weather and sea conditions from weather buoys. While this information has allowed us to develop structural-health-monitoring tools to assess the residual lifetime of the foundations — a research task that is still in process — it has generated a data tsunami involving a lot of different data sources, all with their own format, sampling speed, size and complexity

requiring a scalable and automated data handling.

On the face of it, this looks like a gold mine for researchers but, in practice, a great deal of time must be spent on sorting the large volumes of information and processing it into data streams that can be used to bring new understanding of certain physical phenomena.

Industry benefit

To fulfil our research, OWI-Lab needed a scalable "data factory" that would clean and process the different data streams as they came in, and store them so that the researchers could really focus on the research topics.

For drivetrain monitoring, we wanted to store and process high frequency sampled data instead of only 10-minute values or slow-sampled sensors.

The need to store and handle the large amount of high-frequency sampled data, at sample rates of about 5kHz, is a challenge. Failing to find a solution that could store large volumes of data from different formats at either low or high frequencies, and could be accessed quickly, we developed our own data architecture based on the latest hardware with the help of industry and monitoring experts at the University of Brussels. We financed it with support from the Flemish funding agency for research and technology, IWT.

We will soon be able to use our "hybrid data factory architecture", combining a traditional data warehouse with a state-of-the-art big-data storage and processing approach. We also incorporated approaches from the financial banking sector with expertise on condition monitoring, lifetime assessment, data-mining and the associated wind-turbine application knowledge to create a unique ability to automate the handling of our data.

This is a great step forward for our research, and will help our work to validate new sensor prototypes and innovative condition monitoring approaches for root causes of failures, using greater amounts of data, now that the process is automated.

Although most wind-farm operators have access to

and store certain Scada parameters, many are unable to examine it closely. The data is often not stored for long or becomes dead data unless data scientists are engaged to mine it. Having data is not enough, it needs to be converted into a usable format that will help to cut costs.

Pick your topics

Our experience shows that it is better to gather copious data on one subject, than spread the data thinly across a number of topics. With sufficient historical data interpretation on selected topics, one could, for example, establish if lifetime-extension is possible for an ageing wind turbine, or component.

Such a data project starts with clarification of the failure type or uncertainty to be analysed, and identification of the influencing factors. A list of data sets needed to understand the physical behaviour — and the frequency of samples — can then be generated.

As well as existing Scada data, the captured parameters may also include certain ambient data, such as wind speed, temperature, wave height and humidity levels; even maintenance records could add value. Additional sensors can be added to supply data for any parameters that are missing.

We apply a similar methodology when monitoring the health of offshore foundations, combining data from serval weather buoys and wave radars, certain Scada data that link structural influences, and then we add our own structural measurements on the foundations.

Combining these sets and applying the right data-mining toolsets can lead to magical insights of high value, that will help to support data driven decisions, something that will looks set to grow in importance in coming years. ▪w

Based on historical data interpretation one could examine if lifetime extension is possible for a particular turbine

Pieter Jan Jordaens is business development & innovation manager and Jan Helsen is coordinator for drivetrain monitoring at OWI-Lab. They are speaking at the Drivetrain Component Reliability and Optimisation Forum, 15-16 March in London, UK

CASE STUDY USING FIELD DATA TO MONITOR BEARING WEAR

Field data can be used to monitor bearings through advanced temperature analysis. Variability in the turbine operating conditions means that bearing temperatures fluctuate significantly. This poses a challenge as false positive alarms need to be avoided to keep failure detection effective.

A fast turbine start-up can cause a temperature spike, so the temperature thresholds need to be set well above the highest spike expected during normal operation, to avoid an alarm being triggered in such an event.

The operating regime resulting in the highest temperature will

determine the alarm level, even though it may rarely occur. As such, the failure needs to be sufficiently different to allow the alarm to be triggered.

Combining knowledge of both the physical system and the data can be used to construct an advanced simulation model of normal bearing temperature behaviour. This model can then be combined with anomaly detection to find early-failure.

Physical insights to construct the optimal model can be learned from the available historical data. A high-quality reference data set is extracted from the large data set, representing healthy system

behaviour. It is preferred to extract a separate reference data set for each turbine in the project. The modelled signal is used to predict normal temperature behaviour.

Inputs are measured streams of operational parameters, such as the power produced by the turbine. The temperature signal resulting from the model is compared to the measured temperature signal.

Comparing to the norm

The comparison between both signals characterises the deviation between the expected and the measured temperature

behaviour, revealing the moments of abnormal bearing conditions.

Most of the time, the bearing behaves close to the prediction of the system model and the signal remains within the normal behaviour zone. On several occasions, however, the bearing becomes overheated and the signal enters the warning behaviour zone and even goes over the alarm threshold.

The occurrence and severity of these abnormal events are registered and used to create failure occurrence insights. Based on the degradation accumulation, the operator can decide when to replace a bearing.