DATA PILOT PROGRAMME 2018/19 REPORT

AUTHOR // Conaill Soraghan
DATE // April 2019

In partnership with

PN000290_RPT_006 - Draft 1.5 – Data Pilot Programme 2018-19 Report

Document History

Revision	Date	Prepared by	Checked by	Approved by	Revision History
V1	29/03/2019	Conaill Soraghan	Alistair Lee	Alan Wardlaw	

Executive Summary

Programme Objectives

The Crown Estate (TCE) have commissioned the Offshore Renewable Energy (ORE) Catapult to conduct a series of Data Pilots to help data owners kick-start innovation through exploring data and developing proof of concepts.

Each Data Pilot involves an industrial partner providing a specific data and digital challenge to the ORE Catapult for their data and digitalisation team to tackle by deploying modern digital techniques. This reveals direct value for the industrial partner who receives deliverables such as prototype applications, scripts and interactive dashboards to solve their prioritised problem statement.

The wider purpose of this project is to capture and understand common data and digitalisation challenges facing the offshore wind sector in order to gather industry recommendations.

The Data Pilots

The five Data Pilots delivered in the 2018/19 financial year are:

1. Machine Learning to Predict Power and Identify Leading Edge Erosion
    - How can machine learning approaches be used to identify and quantify leading edge erosion of wind turbine blades from raw SCADA data?
2. Interactive Reporting from 3D Bathymetry Datasets
    - How can bathymetry survey datasets be provided to customers in an interactive way, enabling advanced analysis of the datasets in an accessible format such as a PDF?
3. Hydraulic Pressure System Fault Assessment
    - Can hydraulic oil pressure faults be identified and classified based on signatures in 10-minute SCADA data?
4. SOV and CTV Movement Data Analysis
    - Using AIS data to establish a way to map SOV and CTV vessel movements in relation to wind turbines to identify turbine visits and deduce key maintenance information.
5. Turning Turbine Alarms into Actionable Insights
    - Exploring the use of alarm logs in an operational setting by exploring the available data and verifying a hypothesis on time-based relationships between failures.

Recommendations

The key recommendations that have emerged are:

- **General industry recommendations**
    - The Data Pilot Programme helped many organisations get real value from raw data and the programme could be extended.
    - Barriers exist limiting data sharing and data owners are best placed to take the lead in enabling data sharing and open collaboration.
    - Wind farm raw data sets are large and poorly formatted which requires experts to extract full value and maximise its use. Data owners such as wind farm owner/operators should be encouraged to hire or training up data engineers and data scientists. ORE Catapult are very well placed to provide digital upskilling for the offshore wind industry.
    - A knowledge sharing forum would be beneficial for the sharing of insight and application of innovative data and digital analysis methods. The Crown Estate endorsement would encourage engagement and keep TCE at the forefront of this critical topic.
    - Machine learning has huge potential in the world of data analysis but is not well understood in the wind industry. There is a need for the publication of easily understandable reports and case-studies to educate the industry of this potential.
    - Data owners and service providers who process large and complex data sets will get more value from their data analytics outputs if they consider the capability of the end-user at a very early stage.
    - Automation of repeated data analysis processes is a significant cost saver, but it requires investment in the form of detailed knowledge of underlying data and time to develop useful scripts.

- **Specific technical areas for further research**
    - Machine learning can deliver power prediction and leading edge erosion identification
    - 3D bathymetry interpolation and prediction would be valuable for offshore wind sites
    - Hydraulic pressure system fault prediction is possible with 10-minute SCADA data
    - AIS data is publicly available and is useful for informing offshore wind farm operational KPIs
    - Turbine alarm data is valuable and readily available to asset owners but its use is limited due to difficult formats and lack of understanding. Guidance and tools are required.

Contents

1 Introduction

This report describes the individual Data Pilots that have been commissioned by The Crown Estate (TCE) and delivered by the Offshore Renewable Energy (ORE) Catapult throughout 2018/19.

The main purpose of the report is to explain the problem statements which have been addressed and gather common digital challenges and industry recommendations.

1.1 Data Pilot Programme Objectives

Background/Opportunity

It has been recognised by the ORE Catapult and TCE that data owners in the offshore wind industry such as site owners, operators and manufacturers are struggling with data volume and not fully exploiting modern digital technologies. There is significant potential for operational efficiency gains and cost reduction underpinned by innovations in data use and digital systems.

The barriers to getting value from the operational data are typically either resource/capability bottlenecks or an inability for larger organisations to innovate and try new approaches/processes.

Project Objectives

The Data Pilots Project has been initiated with two key objectives:

1. Help data owners kick-start innovation through exploring data and developing proof of concepts
2. Capture and understand common data and digitalisation challenges facing the offshore wind sector

1.2 A Typical Data Pilot

An individual Data Pilot involves an industrial partner providing a specific data and digital challenge to the ORE Catapult for their data and digitalisation team to tackle by deploying modern digital techniques. This reveals direct value for the industrial partner who receives deliverables such as prototype applications, scripts and interactive dashboards to solve their prioritised problem statement.

A specific Data Pilot takes the following format:

- Innovation workshop between ORE Catapult and industrial partner to explore data and digital challenges and identify a relevant and prioritised problem statement
- Industrial partner shares a sufficient subset of operational data to explore the problem statement(s) identified
- Two months for ORE Catapult data and digitalisation team to develop proof of concepts
- Workshop with industrial partner and ORE Catapult to present proof of concepts and seek industry recommendations

By delivering a set of Data Pilots, the ORE Catapult have detailed appreciation of key data and digital challenges facing a representative sample of the offshore wind sector. We have gathered insight on what the most significant issues are for the industry and where there is commonality across the industrial partners. With this informed view, the ORE Catapult have set out industry recommendations in the final section of this report.

1.3 ORE Catapult

ORE Catapult is the UK's flagship technology innovation and research centre for offshore wind, wave and tidal energy. An independent, not-for-profit business that exists to accelerate the development of offshore wind, wave and tidal technologies. The team of over 150 people has extensive technical and research capabilities, industry experience and track record.

Through world-class testing, research advisory projects and programmes, ORE Catapult work for industry, academia and government to enhance knowledge and improve technology, directly impacting upon the cost of offshore renewable energy.

This programme has been delivered by the Data and Digitalisation Team within the ORE Catapult. This team comprise information system experts, data scientists, mechanical engineers, naval architects and project engineers. The team seek digital transformation projects and opportunities that will either enhance the digital maturity of the renewables industry or enrich the ORE Catapult knowledge and capability in the field of digital transformation.

1.4 The Problem Statements

Throughout the 2018/19 financial year, the ORE Catapult raised awareness of this programme and received many proposals from data owners to participate. Given the resource available, the ORE Catapult were able to deliver five Data Pilots and there remains a pipeline of additional opportunities for further Data Pilots.

The five Data Pilots delivered this year are summarised in Table 1.

Table 1: Summary of 2018/19 Data Pilots

#	Title	Industrial Partner	Problem Statement	Deliverables
1	Machine Learning to Predict Power and Identify Leading Edge Erosion	Offshore wind owner/operator	How can machine learning approaches be used to identify and quantify leading edge erosion of wind turbine blades from raw SCADA data?	- Neural network in the form of a Python script - H5 Datasets - Interactive dashboard
2	Interactive Reporting from 3D Bathymetry Datasets	Marine Surveyor	How can bathymetry survey datasets be provided to customers in an interactive way, enabling advanced analysis of	- 3D PDFs - Word Template files for hosting the 3D images

#	Title	Industrial Partner	Problem Statement	Deliverables
			the datasets in an accessible format such as a PDF?	- Binary and Polygon files to reveal process steps - Matlab code for the creation of 3D animations
3	Hydraulic Pressure System Fault Assessment	Offshore wind owner/operator	Can hydraulic oil pressure faults be identified and classified based on signatures in 10-minute SCADA data?	- Interactive dashboard - SQL queries - Python scripts to identify faults
4	SOV and CTV Movement Data Analysis	Offshore wind O&M advisory service	Use AIS data to establish a way to map SOV and CTV vessel movements in relation to wind turbines to identify turbine visits and deduce key maintenance information.	- Interactive dashboard - Python scripts to calculate KPIs - Excel and CSVs with processed data
5	Turning Turbine Alarms into Actionable Insights	Offshore wind owner/operator	Exploring the use of alarm logs in an operational setting by exploring the available data and verifying a hypothesis on time-based relationships between failures.	- Excel file with results of processing of alarm logs - Interactive dashboard

2 Data Pilots Delivered

2.1 Data Pilot 1 - Machine learning to Predict Power and Identify Leading Edge Erosion

2.1.1 Problem Statement

How can machine learning approaches be used to identify and quantify leading edge erosion of wind turbine blades from raw SCADA data?

To date, industry approaches in the quantification of leading edge erosion impact on performance have included resource intensive lidar campaigns or power performance assessments using unreliable data sources, such as anemometry windspeed data. There is a need for a repeatable methodology to identify leading edge erosion using readily available data streams such as SCADA. Machine learning, as an innovative approach to data analysis, could provide this methodology.

2.1.2 Industrial Partner

The industrial partner is an energy utility company who own and operate onshore and offshore windfarms.

2.1.3 Data

Over three years, 10-minute SCADA data, consisting of over 500 signals, was provided for all WTGs at an operational windfarm. The data included signals regarding the operation, component rotation, rotational speeds and power generation for each WTG.

Over 100 inspection and repair reports were also provided, detailing the erosion level at a number of turbines over the three-year timeframe. Additional data regarding turbine locations, power curves and identification tags were also provided.

Technical meetings between ORE Catapult and the industrial partner enabled knowledge transfer regarding site operation, data availability and detailed understanding of individual signals.

2.1.4 Methods/Techniques

All SCADA data was onboarded, cleaned and reformatted prior to analysis using the Python programming language. The datasets were packaged into hierarchical data formats to enable the analysis of the entire dataset – which would otherwise have been very challenging.

Processing and cataloguing the inspection and repair reports was particularly challenging. As there is not a standardised approach for classifying and categorising erosion damage, the reports were manually reviewed and simple categories (no erosion, slight, moderate and severe) were used to identify turbines which were damaged due to erosion. These turbines were then used as the use-cases for analysis using machine learning algorithms.

An artificial neural network model was created for use in the analysis in order to learn the relationship between target turbines (which were eroded) and its neighbours (which were not eroded). The algorithm takes input signals (active power for several reference turbines) and splits it into several layers which learn discrete behaviours of the relationship between the target and reference/neighbouring turbines. An example of this process is visualised below, showing the hidden layers of the neural network algorithm linking to each other to learn these dependencies.

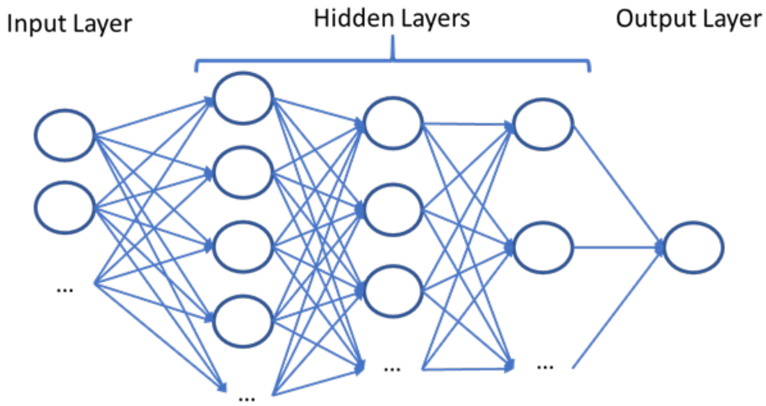

Figure 1: Conceptual architecture of a deep neural network

This model was trained on data after a repair and tested on data before a repair. The change in the turbine's behaviour (from the repair of blade erosion) should then increase the error of the model, indicating a change in turbine performance. The model was created so it could be continuously re-run to identify underperformance due to blade erosion over the entire lifetime of the windfarm.

2.1.5 Results/Outputs

The neural network algorithm provided very promising results. The model itself managed to predict the operation of turbines with a greater degree of accuracy than an averaging methodology – where the active power of a turbine is predicted by taking the average of all neighbouring turbines. This is an industry validated methodology typically used to predict the power of a turbine during periods of data loss.

When the model was then applied to a specific use case, the hypothesised outcome was realised. The results, provided below, show the neural network model's prediction error for data pre and post-repair for six turbines. Erosion was evident on three of these turbines (T01, T02 and T03) and not for the remaining turbines (T04, T05 or T06). All of these turbines were repaired to address the identification of blade structural damage or erosion.

The results show that significant spikes in prediction error are evident on the eroded turbines (T01, T02 and T03), highlighting that the model is now not as accurate due to the blade repair. Conversely, the results for the turbines showing no erosion (T03, T04 and T05) are very consistent pre and post-repair, indicating no performance change due to repair.

This supports machine learning approaches as a valid method to identify leading edge erosion and potentially as a means to quantify its impact on turbine performance.

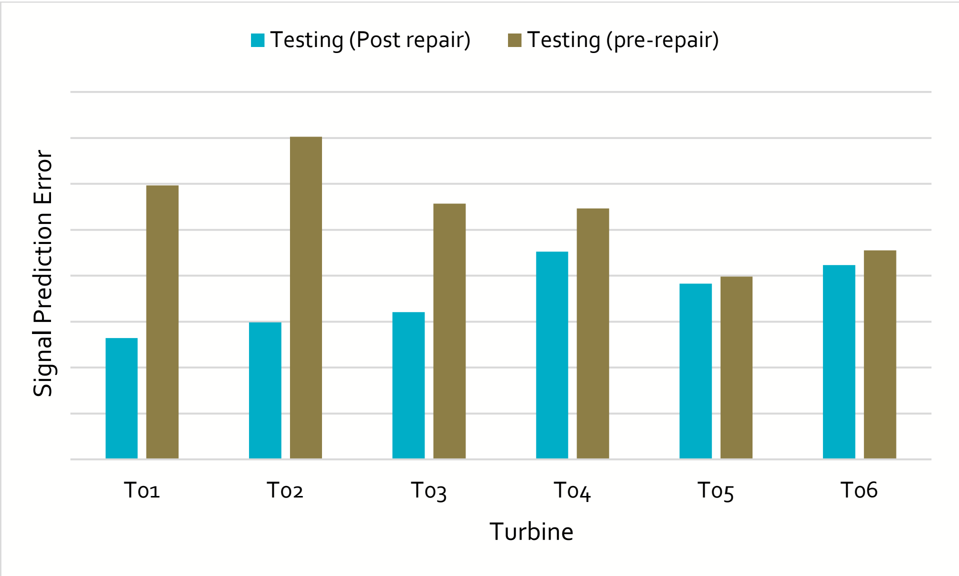

Figure 2: Indicative results of the neural network approach to prediction wind turbine production

2.1.6 Industry Recommendations

- Inspection and repair report consistency is lacking. There is inconsistency in the terminology, image resolution and reporting requirements between inspection/repair service providers; making the evaluation of turbine condition very challenging. It is recommended that erosion classification standards and/or minimum requirement guidelines are produced to provide a common methodology for the industry to adhere to.

- Useful information can be extracted from inspection and repair reports. It is recommended that sites with vast quantities of historical inspection/repair reports go through a manual process of classifying the damage of each turbine. This could be updated with future inspection/repair activity to identify clusters of badly eroded turbines to analyse further.

- Although there are localised teams of data engineers/scientists within many of the organisations operating in the wind industry, there is a lack of collaboration between them. Sharing of insight and application of innovative analysis methods, such as machine learning applications, should be encouraged and actively pursued as the industry develops.

- Machine learning has huge potential in the world of data analysis and is not well understood in the wind industry. There is a need for the publication of easily understandable reports/case-studies to educate the industry around the capabilities and application of machine learning algorithms. This would encourage data owners to fund research project, such as the one described in this report, to further enhance capability and knowledge in this area.

2.2 Data Pilot 2 - Interactive Reporting from 3D Bathymetry Datasets

2.2.1 Problem Statement

How can bathymetry survey datasets be provided to customers in an interactive way, enabling advanced analysis of the datasets in an accessible format such as a PDF?

Currently there is a disconnect between the way survey specialists visualise bathymetry survey datasets and the outputs that are provided to the end customer. There is a need for an accessible way to replicate/imitate the 3D nature of these scans and to provide some means of interactive analysis.

2.2.2 Industrial Partner

The industrial partner is a well-established survey provider operating in the offshore wind industry.

2.2.3 Data

Several point cloud datasets (spatial measurements of the surface of an object/area) were provided as part of this project, these included seabed surveys (such as export cable route scans) and individual object surveys. Object listings (tabulated data recording the objects identified in a survey, such as boulders) were also provided.

As a means to show the industrial partner's current progress with point cloud analysis and survey reporting, example reports, mimicking the reports provided as part of the industrial partner's service offering, were provided.

2.2.4 Methods/Techniques

The key aim of this pilot project was to take raw point cloud datasets and provide 3D PDFs which could be provided as a repeatable service offering by the industrial partner. This involved several steps to convert the raw data into interactive outputs.

Each point cloud dataset was processed using 3D point cloud processing software. This enabled cleaning of the scans (exclusion of outlying points and trimming of boundaries) and analysis of the size and form of the scan itself. An example of a cable exposure from a seabed cable route survey is provided below:

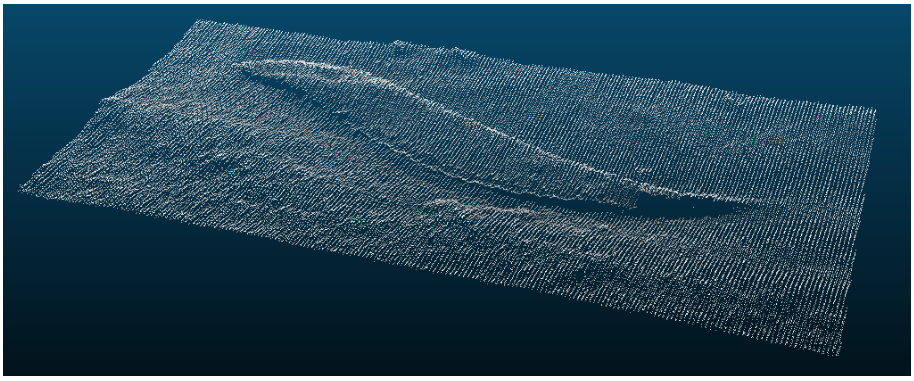

Figure 3: Point could data of an exposed inter-array cable

This 3D point cloud processing software also enabled meshing (3D reconstruction from the surface of a point cloud) which was used for further processing. There are a range of meshing processes which produce varying levels of resolution, storage requirements and detail capture. For this project a Poisson Surface Reconstruction method was utilised to enable high-resolution meshes, which are robust to noise and provide smooth outputs, preferable for the analysis of seabed surveys as in this case. An example of a mesh produced from the seabed cable exposure point cloud is provided below:

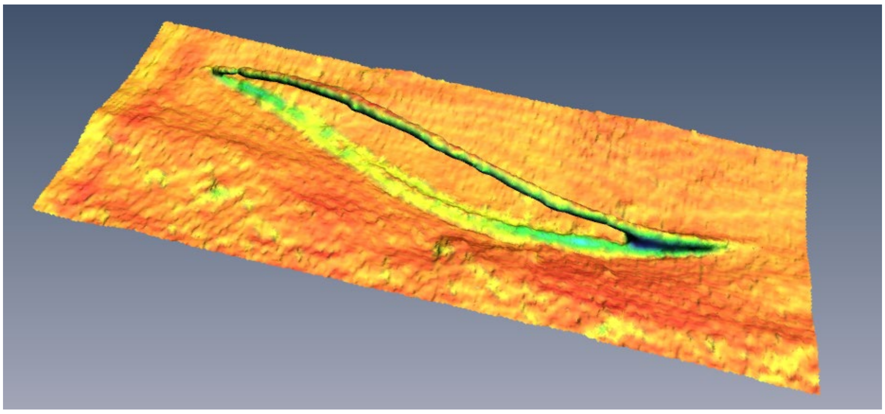

Figure 4: Output of meshing point clouds for an exposed inter-array cable

Meshes were created for several point clouds, then processed and analysed using interactive visualisation software. This enabled bespoke animations to be created and embedded into reports; showing animations of seabed scour and sediment change over time, cross sections of individual objects and other unique visualisations.

Parallel to this analysis, the static object listings were imported into business intelligence software. This allowed for these tabulated datasets to be plotted as interactive maps, allowing the dashboard user to filter datasets based on a number of parameters (such as object size and cable route) and provided an easy way to digest and make sense of complex datasets.

2.2.5 Results/Outputs

Once processing and visualisation was complete, bespoke 3D reports were generated using a 3D PDF generation software. These were provided as examples as to how the complex nature of point cloud datasets can be provided to survey customers as easily-accessible, interactive outputs.

Similarly, the business intelligence dashboards were provided to show how static datasets can be transformed into interactive outputs. These outputs are expected to be helpful in the advancement of capabilities of the industrial partner and as an example of innovative reporting methods for the industry in general.

2.2.6 Industry Recommendations

- The size of raw data/point-clouds is a limitation to analysis and production of outputs. There is a requirement for particular skillsets (such as data engineering and architecture) to be available to organisations with access to these complex datasets.

- The problem statement of this pilot project highlights the need for more interactive and visual means of reporting. There is a disparity between what data specialists/analysts see in their sophisticated software packages and what is delivered to the end customer. The industry needs to think outside of the box to provide outputs in a format that can be easily consumed and understood.
- Automation is key when it comes to processing data. There are a number of complex steps in many data-related tasks which can benefit from automation. This requires a detailed knowledge of the data itself, capability in the area of scripting and automation and a pro-active mindset to enable organisations, and the wider industry, to streamline their processes.

2.3 Data Pilot 3 - Hydraulic Pressure System Fault Assessment

2.3.1 Problem Statement

Can offshore wind turbine hydraulic oil pressure faults be identified and classified based on signatures in 10-minute SCADA data? In particular, three distinct hydraulic pitch and rotor brake system fault types have been observed at the site and the objective of the Data Pilot is to explore if algorithms can be trained to learn from previous periods leading up to these faults in order to provide an indicator of system health and a likelihood of failure of these systems.

2.3.2 Industrial Partner

The industrial partner is an energy utility company who own and operator on and offshore wind farms globally.

2.3.3 Data

A three-year history of 10-minute SCADA data was provided for all WTGs at one operational offshore wind farm. The signals included: Active Power (Avg/Min/Max), Wind Speed (Avg/Min/Max), Rotor Speed (Avg/Min/Max), Pitch (Avg/Min/Max), Rotor Brake Hydraulic Oil Pressure (Avg/Min/Max) and Pitch Hydraulic Oil Pressure (Avg/Min/Max). Furthermore, some static site information was provided including turbine latitude and longitudes and a technical description of the fault types observed at sites.

Historic turbine alarms were also provided for the site but significant challenges prevented this data source being used to add value.

2.3.4 Methods/Techniques

Significant data cleansing and reformatting was necessary before useful analytics could be carried out. SQL was used to hold all of the cleaned data.

Subsequently Python scripts were created to derive further useful metrics and information. For each 10-minute time stamp, operative states and KPIs such as capacity factor and lost production were determined. Another script created a log of hydraulic pressure faults by grouping consecutive 10-minute periods with very low hydraulic pressures.

All of the processed data was combined in a business intelligence dashboard with interactive visualisations so the results of the investigation can be replicated and extended by the data owner. This tool enables identification of the normal operation modes of hydraulic pressure systems and mining the data to find anomalies and indicators of relevant faults.

2.3.5 Results/Outputs

Operational of the hydraulic pressure systems was shown to have two normal modes depending on the power being drawn by the generator. When in low power mode the hydraulic pressure rapidly fluctuates between 180 and 197 kPa. When in high power load, there is tighter control on the pressure. The upper limit of pressure remains the same, but the minimum pressure limit increases to approximately 190 kPa.

The purpose of this investigation is to inform algorithmic detection of deviation from these normal modes of operation. One key finding is that each turbine has its own equilibrium for these control limits.

A hydraulic pressure fault log has been created that reveals which faults are the most significant and which turbines are the worst performers in terms of hydraulic pressure fault downtime. This provided direction for which faults to investigate and is the starting point for creating a list of relevant faults to be used in further analysis such as training machine learning algorithms.

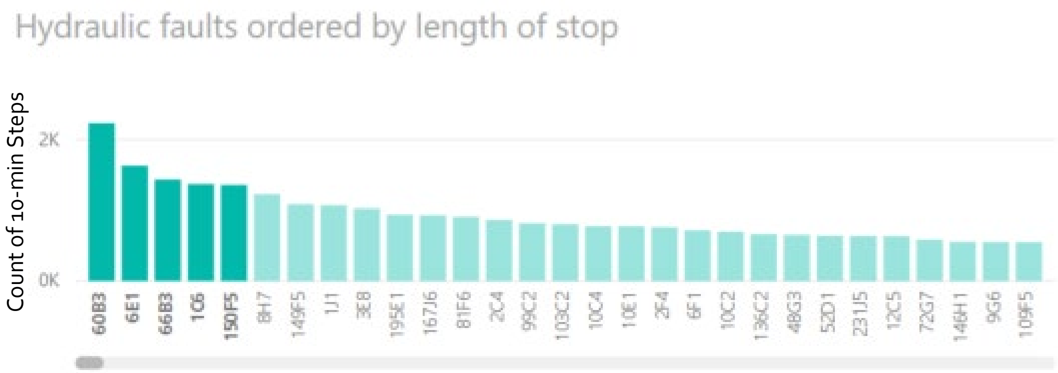

Figure 5: Hydraulic pressure system faults ordered by length of downtime

Finally examples of three distinct hydraulic pressure faults were collected to gather intelligence on what features can be used to identify and classify these faults. The findings are summarised here:

Table 2: Description of three different hydraulic pressure faults

Type	Description	Indicators using 10-min data statistics
Type 1	Causes poor control of pressure at high active power (load) and many trips at high load.	Min hydraulic pressure will be low at high active power. Series of short trips before fault impacts.
Type 2	Control not possible so staging between pressure limits is minimal.	Hydraulic pressure variance (max – min) is very low at any active power
Type 3	Pressure decays before a shutdown occurs. This trend often repeats on restarts.	Max hydraulic pressure decays over a 10-30 minute period

2.3.6 Industry Recommendations

- Hydraulic pressure faults are causing significant downtime at operational sites and there are signatures in 10-minute SCADA data that could be exploited to automatically identify and classify these faults.
- Better definition of data sharing agreements, sanctioned data sets to be shared and specific problem statements (requirements) are necessary to enable effective cross industrial collaboration.
- Turbine alarms could not be used despite being available for the Data Pilot. Standardisation and better description of this data source is crucial to help data owners get value form this source.
- Raw data sets are large and poorly formatted which requires experts. Data owners such as wind farm owner/operators should be encouraged to hire or train up data engineers and data scientists.

2.4 Data Pilot 4 - SOV and CTV Movement Data Analysis

2.4.1 Problem Statement

Establish a way to map SOV and CTV vessel movements in relation to wind turbines to identify turbine visits and deduce key maintenance information. Use this information to:

- Establish how SOVs are being used at operational windfarms
- Identify the value of a supporting Crew Transfer Vessel (CTV)
- Establish key maintenance metrics and KPIs using publicly available data
- Understand what changes to the working data/routine are introduced when an SOV is used

2.4.2 Industrial Partner

The Industrial Partner is a consultancy that specialises in Operations and Maintenance (O&M) support for renewable energy projects. This company undertakes varied work in renewables O&M; from asset management solutions to undertaking strategic projects and carrying out performance improvement activities in support of their client base.

2.4.3 Data

AIS data was purchased from the publicly available website Vessel Finder. Five-minute time period historical data was purchased for two SOVs and one CTV:

- Esvagt Njord (SOV) – Dudgeon
- Dalby Ouse (CTV) – Dudgeon
- Edda Passat (SOV) – Race Bank

Wind turbine location details were also acquired from a publicly available data source. Seafish produce 'Kingfisher Wind Farm Charts' for mariner awareness, these are available for free download from KIS-ORCA. Turbine, substation and buoy locations were obtained for Dudgeon and Race Bank.

2.4.4 Methods/Techniques

The raw data was manipulated using Python, using the modules Shapely and Pyproj. These modules allowed latitude and longitude data to be transformed into Python points, allowing coordinates to be compared.

A Python script was developed to determine when a vessel was near an asset, how close it was to the asset and what asset it was. This data was then stored in a number of excel sheets to be read by the PowerBI file. Namely a sheet with all timestamps and if a vessel was near an asset and a sheet with all visiting events, detailing the start and end of the event as well as the average visiting distance and length of visit.

A PowerBI dashboard was created to take these scripts and show them in a user-friendly manner. The final dashboard allows a user to interact with graphs and lets them slice the data to draw their own conclusions.

2.4.5 Results/Outputs

It was found that the vessels work in very different manners. For example, time spent at a turbine varied greatly between the SOVs and the CTV, with the CTV spending much longer periods near a turbine. Differences were also found between the two SOVs, Esvagt spends a much greater length of time near a substation when compared to Edda, see Figure 6 (right).

Looking closer at the length of time each vessel visits a turbine it can be seen that there is a large difference between all vessels. The SOVs are tighter grouped, with most visits being below 200 minutes whereas the CTV is much more spread, with visits going over 500 minutes long. Another observation that can be made from the histogram is the location of the spike. The Edda has a spike around 20 minutes, corresponding to the average time for a technician transfer, but the Esvagt has a more varied spike, ranging from 20 minutes to 40 minutes, showing longer transfers. The CTV's transfers are much shorter, with the spike in transfers occurring at 5 to 10 minutes, this will correspond to the CTV coming alongside the asset to quickly drop off technicians.

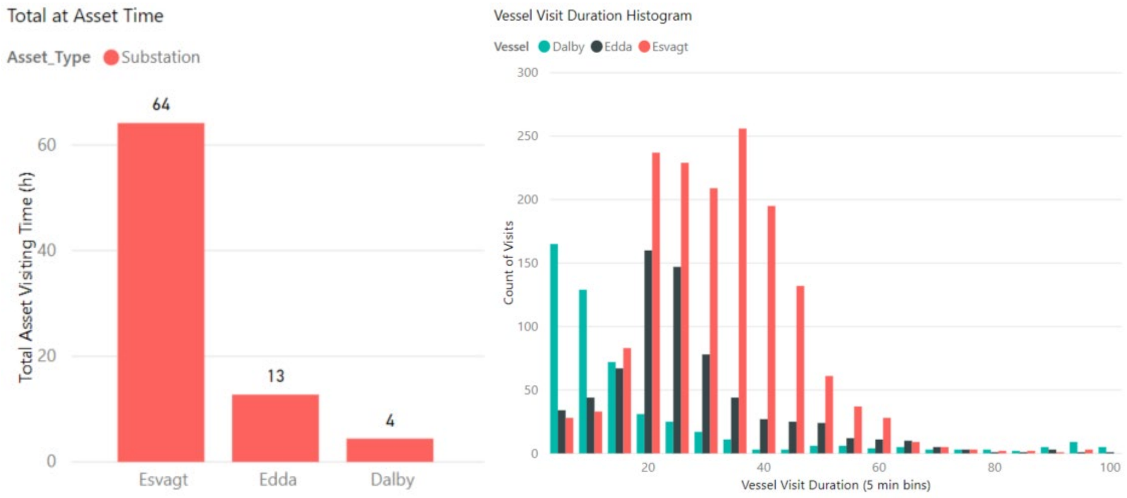

Figure 6 - Total Time at Substation (left) & Vessel Visit Duration Histogram (right)

As well as this report the SME was provided with the Python script used to perform the data manipulation, the PowerBI file used to draw conclusions and the full list of vessel visit events.

Overall the project can be called a proof of concept; AIS data can be used to get an understanding of how vessels are being used at an operational wind farm. It was not possible to prove the effectiveness or otherwise of a supporting CTV, this was due to lack of analysis time, although this approach appears to have the potential to do this. The data can also be used to deduce high level maintenance information which offers an opportunity for benchmarking and also has potential to provide an open source of O&M information to third parties. There is a recognised lack of publicly available O&M data for operational offshore windfarms and the use of AIS ship movements offers a means to obtain some information. This will not have the richness of full-blown maintenance history and inspection reports but it could provide simple metrics including visits to turbines, duration of visits and maintenance frequency.

2.4.6 Industry Recommendations

- The Crown Estate should encourage the use of AIS and other publicly available data for increased learning. Owner/operators could learn more about their vessels and how best to optimise, or this information could be used by SMEs and academia to create new innovative solutions for industry, for example:
  - Optimisation of maintenance routes
  - Linking with other publicly available data sources to further utilise AIS data
  - Optimisation of just using an SOV or using an SOV and CTV together

- Industry should clearly state what KPIs could be of use for increasing productivity at site. Solution providers, such as SMEs, should know what KPIs they should be targeting and why, this will help concentrate their work to the correct area.
- Better integration between O&M specialists and data specialists should be encouraged. O&M specialists know what is happening at site but don't know how to use the emerging data sources to strengthen their knowledge base. Data specialist know how to exploit data techniques but don't know the domain area well. Integration between these two would help strengthen the industry and lead to good collaboration and innovative ideas forming.
- The process for obtaining AIS data was simple. The website was easy to contact and when data was downloaded it came in a clean format that was well documented. Processes such as this should be encouraged by other data providers to allow publicly available data to be easily used.

In the future there is room for much more work, for example looking at work routes, how are the vessels going between turbines, in a step-wise manner or going around the wind farm in a less choreographed manner.

2.5 Data Pilot 5 - Turning turbine alarms into actionable Insights

2.5.1 Problem Statement

Analysing alarm logs in a systematic way to deriving insights in, for example, component failure rates, is extremely challenging but has the potential to underpin significant benefits for performance and reliability analysis.

In this data pilot we are exploring the first steps of using alarm logs in an operational setting by exploring the available data and trying to verify a complicated hypothesis on time-based relationships between failures: are there clear cases where a long lasting outage is preceded by several short ones see Figure 7.

Figure 7: short outages followed by a long outage

2.5.2 Industrial Partner

Operator of a large windfarm.

2.5.3 Data

For this data pilot, raw and processed alarms are shared. The raw alarms came straight from the turbine and the processed ones included alarm descriptions, stop type and component categories for only alarms that stop the turbine. For these alarms, clusters are defined for all strictly overlapping alarms.

2.5.4 Methods/Techniques

The main analysis that is performed tries to verify the hypothesis of the existence of specific sequences of alarms. I.e. the scenario were several short failures are followed by a longer failure.

To do so, failures are defined, and the two most suited components are picked as examples. For these components, a follower relation is defined, calculated and visualised. Using Cumulative Density Functions on the duration between failures, a relation between shorter and longer alarms is found.

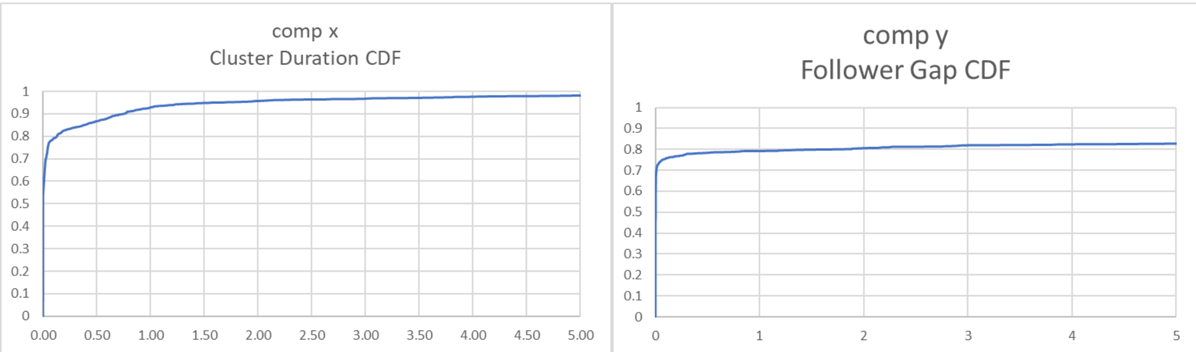

Figure 8: CDF on outage duration and time between two consecutive outages.

2.5.5 Results/Outputs

Sharing the data was technically easy to accomplish but receiving approval from all needed stakeholders was the most time-consuming part of this pilot. When the data was received, it was great to see that the owner had already taken steps to enrich the alarm logs by adding information on the stop type and component of failure. Overlapping alarms were also already grouped into clusters, which made it easier to focus on outages.

There are clear scenarios where short failures are followed by a longer failure. However, the time in-between the failures is a cause for caution and needs further refinement by, for example, including knowledge from the operational team of the windfarm.

2.5.6 Industry Recommendations

- Practices should be appraised to enable enhanced data sharing. Even though all parties are aware of the added value of sharing data, and having experts from outside of the business involved, sharing data still remains a significant challenge.

- Alarm data typically comes from the turbines in a raw format that is hard to use in a structural way. Adding information like alarm descriptions, stop type and component, enables structural approaches. However, alarms still relate heavily to one another, for example, due to overlap.

Clustering alarms into groups that belong to one cause adds value because this enables structural failure analyses, as is done in other industries.

- Asset owners should be supported in how to convert raw alarms into actionable maintenance KPIs. In the process of performing the analysis, other potentially interesting insights came to light in the form of higher-level metrics, used in other industries, that can be calculated using the available data. These metrics can be used by operations to see how well they are performing. Examples are illustrated in Figure 9.

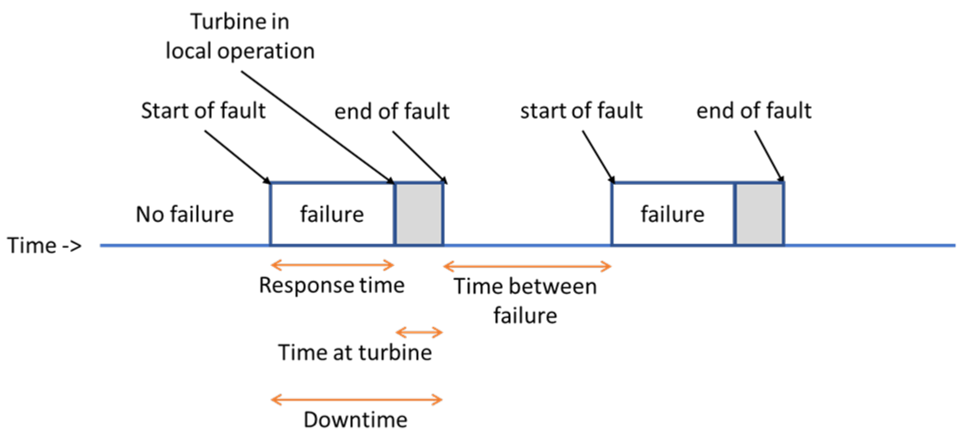

Figure 9: High level KPI possibilities using alarm logs

In general, it is clear that alarm logs have the potential to inform analysts on what has been going on in the turbine, which can be used to find trends and causes.

3 Industry Recommendations

3.1 General industry recommendations

- **The Data Pilot Programme helped many organisations get real value from raw data and the programme could be extended**

The ORE Catapult were oversubscribed in terms of industrial partners and problem statements based on the funding available for the project. Furthermore, there were problem statements that were rejected as they were not directly relevant or too close to other projects such as SPARTA.

Given that there was very little industrial awareness of this programme, this demand for Data Pilots was unexpected. There is now a healthy pipeline of additional industrial partners and problem statements for further Data Pilots going forwards.

There is significant potential to continue this programme in subsequent years.

- **Barriers exist limiting data sharing and Data owners have the ability to take the lead on enabling data sharing and open collaboration**

This programme of work has demonstrated that open collaboration between data owners and solution providers can be extremely effective, but data sharing is necessary. The ORE Catapult delivered all of these Data Pilots internally, but across the industry there is a healthy community of SMEs, researchers and academic organisations extremely keen to solve similar data-driven problems. However, significant barriers exist, limiting open collaboration.

For most of the Data Pilots, there were unexpected delays commencing the work and improving working practices could encourage much more collaboration across the industry. The delays were typically due to industrial partners spending necessary time reviewing data sharing agreements to ensure data owners data and IP is protected.

Furthermore, the scope of many of the Data Pilots altered because the required data could not all be provided, which caused the problem statement to evolve.

Therefore, as a recommendation, to enhance data sharing and cross industry collaboration, Data owners should take the lead on:

- Defining the specific data sharing agreement they want to use. Clearly stating what data is needed, what the purpose of the data sharing is, who retains IP and who would receive the results and outputs.
- Identify, sanction and describe the data that can be shared.
- Clearly state the problem statements that are important to them.

- **Wind farm raw data sets are large and poorly formatted which requires experts to extract full value and maximise its use. Data owners such as wind farm owner/operators should be encouraged to hire or training up data engineers and data scientists. ORE Catapult are very well placed to provide digital upskilling for the offshore wind industry.**

The points of contact from the industrial partners were all experts in the engineering domain relevant to their individual problem statement, however many did not have access to capable data engineers and data scientists within their organisations to prepare data in such a way that it can be readily used for analysis.

Therefore, a large proportion of the time of many of the Data Pilots was spent writing Python scripts to clean and reformat data before it could be loaded into a tool such as SQL that can handle the size of the data sets generated by WTGs. This was only possible with dedicated Data Scientists within the ORE Catapult.

Some of the specific data cleansing tasks involved algorithmically altering corrupt values, redefining formats and pivoting key-value pair structures.

Given the volume, variety and velocity of data generated by offshore wind assets, data owners such as wind farm owner/operators should be encouraged to hire or train up data engineers and data scientists to improve the efficiency of data analysis and increase the proportion of data that is actually used.

The ORE Catapult are very well placed to provide digital upskilling. With a very deep knowledge of the offshore wind digital systems in use and resident computer scientists with data engineering and data science experience, this resource should be available to the industry for training and developing proof of concepts.

- **A knowledge sharing forum would be beneficial for the sharing of insight and application of innovative data and digital analysis methods. The Crown Estate endorsement would encourage engagement and keep TCE at the forefront of this critical topic.**

Although there are localised teams of data engineers/scientists within many of the organisation operating in the wind industry, there is a lack of collaboration between them. Sharing of insight and application of innovative analysis methods, such as machine learning applications, should be encouraged and actively pursued as the industry develops. All participants asked for some form of mechanism to further discuss this critical topic, so the demand is high.

A knowledge sharing forum would be an excellent mechanism for sharing state of the art analytics methods, lessons and insights throughout the offshore wind sector.

The Crown Estate would be an excellent organisation to act as sponsor of such a forum. Not only would this encourage engagement, but the Crown Estate would benefit significantly from the open sharing of digital challenges and pilot studies.

- **Machine learning has huge potential in the world of data analysis but is not well understood in the wind industry. There is a need for the publication of easily understandable reports and case-studies to educate the industry.**

Machine learning has huge potential in the world of data analysis and is not well understood in the wind industry. There is a need for the publication of easily understandable reports/case-studies to educate the industry around the capabilities and application of machine learning algorithms. This would encourage data owners to engage in and fund innovation projects, such as those described in this report, to further capability and knowledge in this area.

User guides like this would be a very useful output of the knowledge sharing forum proposed in the previous recommendation.

- **Data owners and service providers who process large and complex data sets will get more value from their data analytics outputs if they consider the capability of the end-user at a very early stage.**

There is a disparity between what data specialists/analysts see in their sophisticated software packages in-house and what is meaningful and useful for the end customer. This also includes internal customers across an organisation.

Data owners and service providers who process complex data sets are encouraged to consider end-users at the early stages of carrying out data analysis. The industry needs to think outside of the box to provide outputs in a format that can be easily consumed and understood.

- **Automation of repeated data analysis processes is a significant cost saver, but it requires investment in the form of detailed knowledge of underlying data and time to develop useful scripts.**

Automation is key when it comes to processing data. There are a number of complex steps in many data-related tasks which can benefit from automation. This requires a detailed knowledge of the data itself, capability in the area of scripting and automation and a pro-active mindset to enable organisations, and the wider industry, to streamline their processes.

3.2 Specific technical areas for further research

- **Machine learning can deliver power prediction and leading edge erosion identification**

Machine learning has been demonstrated to be an effective approach for predicting power at an operational wind farm. Further research and development will optimise the algorithms to improve accuracy and result in user-friendly tools.

Using ML for more specific use-cases such as identifying leading edge erosion (LEE) has shown potential, but improvement will rely on training on known cases of LEE. Unfortunately, there is inconsistency in the terminology, image resolution and reporting requirements between inspection/repair service providers; making the evaluation of turbine condition very challenging.

It is recommended that erosion classification standards and/or minimum requirement guidelines are produced to provide a common methodology for the industry to adhere to.

- **3D bathymetry interpolation and prediction would be valuable for offshore wind sites**

Bathymetry surveys are critical for offshore wind farm site and asset management, yet they are prohibitively expensive limiting the volume of surveys. One area explored in this programme is the potential to interpolate between static images to reveal the time evolution of bathymetry features. Moreover, using this historic information to determine rates of change can lead to useful predictions of how sites are likely to evolve over the life of the wind farm.

- **Hydraulic pressure system fault prediction is possible with 10-minute SCADA data**

Hydraulic pressure faults are causing significant downtime at operational sites. This investigation has revealed that there are signatures in 10-minute SCADA data that could be exploited to automatically identify and classify these faults. There is significant potential for further research on this topic.

- **AIS data is publicly available and is useful for informing offshore wind farm operational KPIs**

The Crown Estate should encourage the use of AIS and other publicly available data for increased learning throughout the industry. Owner/Operators could learn more about their vessels and how best to optimise, or this information could be used by SMEs and academia to create new innovative solutions for industry. Example topics include, optimisation of maintenance routes; linking with other publicly available data sources to further utilise AIS data; optimisation of just using an SOV or using an SOV and CTV together.

- **Turbine alarm data is valuable and readily available to asset owners but its use is limited due to difficult formats and lack of understanding.**

Turbine alarm data contains potential value as a timestamped log of how the turbine is being controlled. Each alarm has a unique ID number that is linked to a brief text description. While alarm logs were available for three of the Data Pilots, due to a lack of availability and understanding of alarm descriptions, overwhelming volume and bad formatting of the alarm log, it was not possible to use this data source in two of the three relevant Data Pilots. The impression was gained that owner/operators are also reluctant to use this data source for performance analysis for similar reasons.

To resolve this issue there are various efforts that would be of significant industrial benefit.

- Fundamentally, OEMs should endeavour to better describe alarm logs as they design the technology that generates the alarm logs.

- Standardisation of this data source is necessary and a workgroup should be established to work towards standards, generate guidelines on how to interpret alarms and deliver tools that convert the raw alarms into useful processed alarm logs.

- Processing alarms, by for example categorising (e.g. assigning stop types and high level components), and clustering increases usability drastically.

Contact

GLASGOW

Inovo
121 George Street
Glasgow
G1 1RD

T +44 (0)333 004 1400

BLYTH

National Renewable Energy Centre
Offshore House
Albert Street
Blyth, Northumberland
NE24 1LZ

T +44 (0)1670 359 555

LEVENMOUTH

Fife Renewables Innovation Centre (FRIC)
Ajax Way
Leven
KY8 3RS

T +44 (0)1670 359 555

HULL

O&M Centre of Excellence
Ergo Centre
Bridgehead Business Park
Meadow Road, Hessle
HU13 0GD