DNV GL: Multi-Model and Machine Learning Wind Forecast Project

LESSONS LEARNT REPORT 3

Project Details

Recipient Name	DNV GL Australia Pty Ltd (ABN 14 154 635 319)
Contact Names	**Trenton Gilbert**, Head of Section, Developer Support Services **Mark Purcell**, Senior Engineer, Developer Support Services **Michael Quan**, Engineer, Developer Support Services **Graham Slack**, Country Manager, Australia - Energy
Contact Details	Trenton Gilbert: trenton.gilbert@dnvgl.com Mark Purcell: m.purcell@dnvgl.com Michael Quan: michael.quan@dnvgl.com Graham Slack: graham.slack@dnvgl.com
Reporting Period	May 2020 – September 2020

This Project received funding from ARENA as part of ARENA's Advancing Renewables Program - Short Term Forecasting funding round.

EXECUTIVE SUMMARY

DNV GL have entered a funding agreement (the Project) with Australian Renewable Energy Agency (ARENA) to provide short term wind power forecasts for the Ararat Wind Farm. The purpose of the Project is to explore the potential for wind farms to provide their own, more accurate, forecasts as inputs into AEMO's central dispatch system.

DNV GL has partnered with the Ararat Wind Farm Pty Ltd (AWF) and RES Australia Pty Ltd (RES). AWF represent the wind farm for which forecasts are to be provided, and are providing access to relevant data and allowing forecasts to be submitted on behalf of the Project. RES are providing support with facilitation of the Project, and evaluation of the value that accurate forecasts can provide to an operational project.

Two of the main challenges associated with the Project to date have been successful recovery of real-time data from the site, and communication via the AEMO MP5F API. These challenges were also present earlier in the Project [DNV GL Lessons Learnt Report 1] and [DNV GL Lessons Learnt Report 2] and whilst they have largely been overcome, they are still persisting to a degree, and have the potential to impact upon the reliability of forecast delivery. The key lessons learnt are that it is necessary to allocate adequate resources to both aspects of the Project, across all stakeholders involved in the Project, and that it is important to establish processes to pre-emptively prevent outages or resolve them when they occur.

Further challenges have included developing and implementing a machine learning model to generate accurate short horizon forecasts. Considerable effort has been made to fine tune the machine learning parameters while also reducing latency in live site data and model run time.

KEY LEARNINGS

Some of the challenges faced and achievements made by the Project to date have included:

- establishing a process to recover real-time data from the site,
- communicating with the AEMO API and
- developing an accurate machine learning model
- observations that improved forecast accuracy does not necessarily lead to improved financial outcomes.

These are discussed further below.

Lesson learnt No. 1:

Category: Technical

Objective: Recovery of real-time site data

The process of recovering real-time data from the site has been challenging. Currently DNV GL has been able to recover real-time site data through the third-party dispatch solution provider, and also directly from AEMO dispatch data. Real-time site data retrieval is currently achieved by establishing a secure connection to third-party dispatch solution provider, which has not been sufficiently robust in some cases. It has not been possible to obtain access to real-time site data through other mechanisms at this stage, primarily due to security restrictions and/or knowledge gaps regarding details of the wind farm SCADA system. Opportunities to increase the fidelity and reliability of the data retrieval are currently being explored, an alternative being to retrieve data via a web application programming interface (API). This may be implemented in later stages of the project to improve the reliability of the data retrieval process, and hence timely delivery of accurate forecasts.

Implications for future projects:

The challenge of recovering real-time data from the site is likely to be unique to each future project. As such, it will be important to allocate adequate resources to this activity for future projects. Extensive discussions with all parties involved in data acquisition and management for the generator are necessary in order to fully understand the environment and establish the most appropriate method for data retrieval.

Lesson learnt No. 2:

Category: Technical

Objective: Connection with AEMO MP5F API
The process of connecting to and interacting with the AEMO MP5F API has been challenging. It has previously been noted that the process could potentially be improved though clearer documentation and access to additional resources to assist with troubleshooting, and was complicated in situations where there are multiple forecast providers.

Currently the main challenges are related to ensuring the robustness of ongoing forecast delivery. DNV GL has experienced an inability to connect successfully to the API at certain times as a result of both credential expiry, certificate revision, and AEMO outages.

The primary lessons learnt to date are that it is important to have processes in place to proactively deal with outages caused by credential expiry or other issues. DNV GL's Forecaster service fortunately provides processes to mitigate the risk associated with these issues, however in some cases (such as AEMO outages) they are outside of DNV GL's control to resolve independently, which has led to unsuccessful delivery of forecasts in some periods.

Implications for future projects:
The challenges associated with establishing communication with the AEMO API and submitting forecasts have largely or will be overcome, and lessons learnt will be applicable to future projects. It is necessary to ensure that the process for interacting with the API is well understood, and that processes are in place for dealing with issues such as credential expiry and outages.

Lesson learnt No. 3:

Category: Technical

Objective: Development of an accurate machine learning model

Current research and developments in machine learning processes has led to several robust models that can be adopted for different forecasting scenarios, where extensive investigation of dependent or independent variables or “features” is required. Features typically considered include variables such as seasonal and diurnal variation, power production, wind speed, wind direction and other atmospheric parameters. Each model is tuned based on different feature sets, ranging from low frequency AEMO dispatch data to high frequency SCADA data. The models built are ensemble models based on gradient boosting machine learning techniques. Features available in raw data streams are put through a process of feature engineering which is investigated concurrently with model development. Feature engineering involves the use of specialised domain knowledge to generate new features from existing features, which are in turn fed into a machine learning model to improve its predictive performance. Each model requires careful tuning of model parameters based on the set of feature engineered inputs to obtain the best result for each situation.

DNV GL has tested a number of different feature engineered input data streams for a range of model setups to develop the current machine learning approach that is being deployed for the Project wind farm. In doing so, DNV GL has learnt how sensitive the forecast accuracy is to different inputs and which models provide a robust prediction that is accurate in most wind conditions. As expected, change points in the wind farm production time series brought on by a sudden changing wind climate are difficult to predict and further work developing the machine learning model will likely focus on these periods.

Implications for future projects:

The work done to develop and tune a machine learning model will provide a good framework for similar short horizon forecasts for future projects. However, each new site will present unique challenges and require unique training of the machine learning model with bespoke settings to provide the most accurate result. Sensitivities of the model accuracy to training data length have not yet been fully investigated, however it is envisaged that multiple months of input data are likely to be required, and one year of data is likely to be desirable to capture variations in production due to seasonality. This however may also be unique to each project, depending on how strongly production is affected by seasonality.

Lesson learnt No. 4:

Category: Commercial

Objective: Evaluation of financial outcomes

Although current developments of machine learning processes have been optimised to lower the relative the Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE), the primary objective of the trials are to deliver improved forecasts in order to ultimately enable significant financial savings for participants. Results from preliminary investigations into the impact of forecast accuracy on FCAS charges have indicated that while improved forecast accuracy can lead to improved outcomes, this is not always the case. As such, improved self-forecasts relative to AWEFS may not necessarily always translate to lower FCAS charges.

Therefore, the opportunity potentially exists to improve outcomes by developing a forecast model which is optimised to have improved forecast accuracy and improved financial outcomes. This line of investigation may be explored in the remainder of the project.

Implications for future projects:
In order to confidently realise benefits from improved forecast accuracy, it may be necessary to also consider development of forecast models which are optimised to minimise FCAS charges. The complexity and dynamic nature of the FCAS market and charges make this a challenging problem.