NIA Project Annual Progress Report Document

Date of Submission

Jul 2022

Project Reference Number

NIA2_NGESO003

Project Progress

Project Title

Probabilistic Machine Learning Solution for Dynamic Reserve Setting

Project Reference Number

NIA2_NGESO003

Funding Licensee(s)

NG ESO - National Grid ESO

Project Start Date

May 2021

Project Duration

2 years and 3 months

Nominated Project Contact(s)

Gabriel Griffin-Booth

Scope

Dynamic Reserve Level Setting Approach

- The different types of reserve setting processes should be integrated to ensure that the total reserve held is representative of NGESO's risk appetite.

- In this project, we expect to apply the approach summarised below to both the basic reserve (possibly including interconnectors) and the reserve for renewable energy sources and we expect that our focus will be on 4–24-hour lead times.

Dynamic Reserve Setting Solution Design

This project will create a proof of concept for a DRS solution with the following features:

- Data pipeline which automates the extraction, cleaning, and preparation of raw data into storage.

- Probabilistic ML (machine learning) model which makes use of predictor variables in that data (e.g. temperature and wind forecast quantiles, and generation mix) to create more accurate predictions (and prediction intervals) of forecast errors and therefore set reserve levels, which better reflect NGESO's risk appetite.

- A database and dashboard for the display of the results of the ML model.

- Automated upload of the ML model results to control room systems.

- Retraining of the ML model to enable a cycle of continuous learning using new data about recent system conditions and forecast errors.

Objectives

- Probabilistic ML model which makes use of predictor variables in that data (e.g. temperature and wind forecast quantiles, and generation mix) to create more accurate predictions (and prediction intervals) of forecast errors and therefore set reserve levels, which better reflect NGESO's risk appetite.

- Retraining of the ML model to enable a cycle of continuous learning using new data about recent system conditions and forecast errors.

Success Criteria

Optimising reserve levels and reducing balancing costs is a key challenge for the NGESO as we move towards our net zero targets and a more volatile system. Uncertainty is more difficult to predict in the short term and we are seeking to optimise reserve levels in the most efficient way.

The end result would be better optimised levels, avoiding 'overholding' and better value for end consumer, and avoiding 'underholding' in risky and uncertain periods.

Performance Compared to the Original Project Aims, Objectives and Success Criteria

National Grid Electricity System Operator ("NGESO") has endeavoured to prepare the published report ("Report") in respect of Probabilistic Machine Learning Solution for Dynamic Reserve Setting, NIA2_NGESO003 ("Project") in a manner which is, as far as possible, objective, using information collected and compiled by NG and its Project partners ("Publishers"). Any intellectual property rights developed in the course of the Project and used in the Report shall be owned by the Publishers (as agreed between NG and the Project partners).

The Report provided is for information only and viewers of the Report should not place any reliance on any of the contents of this Report including (without limitation) any data, recommendations or conclusions and should take all appropriate steps to verify this information before acting upon it and rely on their own information. None of the Publishers nor its affiliated companies make any representations nor give any warranties or undertakings in relation to the content of the Report in relation to the quality, accuracy, completeness or fitness for purpose of such content. To the fullest extent permitted by law, the Publishers shall not be liable howsoever arising (including negligence) in respect of or in relation to any reliance on information contained in the Report

Copyright © National Grid Electricity System Operator 2022

Project Overview

Reserves is a term used to describe spare MWs which are synced ahead of time and are used to correct for changes to the volume of demand and supply between a forecast and real-time. Currently, the volume of reserves which is scheduled on the system is determined by a calculation which is performed bi-annually, once at BST clock-change and once at GMT clock change. This project aimed to develop a Proof of Concept (PoC) probabilistic machine learning model for allocating reserve dynamically based on the latest real-time and forecast system data. Reserve levels must be sized efficiently such that there is always an appropriate volume of reserve on the system. In this definition, appropriate is defined such that there is always enough reserve to operate the system according to NGESOs risk appetite whilst also minimising excess reserve holding where possible, thus minimising costs.

The project specifically looked at applying ML techniques to the problem of forecasting operational reserves. One key requirement from the control room for any reserve's recommendation methodology is that there must be some accompanying explanation for why the model has suggested the reserve levels. Therefore, the Dynamic Reserve Setting project aimed to deliver a set of explainable ML models which use the latest real-time and forecast data to recommend reserve which meets NGESOs risk appetite.

The work for the Dynamic Reserve Setting project can be broken down into two parts: pre-processing the data and creating the skeleton of a data processing pipeline for any reserves forecast and using the processed data to develop and train PoC explainable ML models for forecasting reserves.

Part 1: Data Processing

The first stage of the project involved processing a large volume of internal NGESO data. The data was downloaded from multiple NGESO systems, each data source was cleaned to remove any outliers and erroneous data points, missing data was handled appropriately, every data source was aligned temporally and spatially, and finally a MySQL database was constructed to house all the cleaned data. In total, the data processing consumed two thirds of the total project time, but it was necessary for facilitating the modelling stage of the project and for ensuring quick integration of the project outputs internally within NGESO following successful project completion.

Part 2: Modelling

Once the data was cleaned, processed and stored in the database, the modelling stage of the project commenced. The modelling can

be broken down into four steps:

- constructing the target variable
- performing exploratory data analysis to understand the relationships between the target variable and all available input variables
- building an ML model for forecasting given percentiles of the target variable
- developing a method for explaining how the ML model arrived at each reserve recommendation

Constructing the target variable: reserves are used to provide resilience against all sources of error and uncertainty which affect system operation at some leadtime. To know how much reserve to hold we must first be able to quantify the uncertainty in system operation at any given leadtime. In general, there are four contributions to system level uncertainty which we can easily measure: solar forecast error, wind forecast error, demand forecast error, and power plant unreliability (either changes to Physical Notifications by real-time or under, or over, delivery in real-time or both). Calculation of solar, wind, and demand forecast errors is simply the difference between the predicted and the actual outturn measured in MW. If power plants operated exactly as planned with zero downtime or faults, then the reserve volume would only be dictated by the error and uncertainty in these forecasts. However, power plants deviate from their planned position for numerous reasons. For example, unplanned maintenance, a fault / trip, and over or under delivery of power. Quantifying the error associated with power plant unreliability is not straightforward because any calculation must account for expected changes to power plant position between some leadtime and real-time e.g. changes to power plant position caused by a bid or offer acceptance in the control room.

To quantify the error associated with power plant unreliability two metrics called the upwards reliability error and the downwards reliability error were developed which calculate the volume of headroom and footroom on the system at some leadtime and compare it with the volume at real-time, whilst accounting for faults and control room actions. Appendix A of the project closedown report, available on the Smarter Networks Portal, describes in more detail how the upwards and downwards reliability errors are calculated.

Finally, two target variables were constructed for upwards and downwards reserves by summing the solar, wind, demand, and upwards / downwards reliability errors together. The errors are summed so that increases in headroom/footroom can offset any forecast errors on a per settlement period basis.

Exploratory data analysis: once the two target variables had been created, a phase of exploratory analysis was kicked off. In this stage of work, relationships between the target variables and all input predictors were explored to identify the best features to use in the ML model. Considered input features included all of NGESO's weather forecast data, all demand and generation forecast data, and all settlement and metering data for Balancing Mechanism Units (BMUs). The best identified features are the mean and standard deviation of wind speed forecast across all forecast locations, the national embedded wind load factor forecast, the national embedded PV load factor forecast, the 1-day lagged embedded wind load factor, the 1-day lagged national demand, and a Boolean flag indicating whether the day was a bank holiday or not.

Building the ML model: once the best features were identified the next stage involved building the ML model. Different ML modelling approaches were tested and the python-based Scikit-learn LightGBM framework was selected. This model was selected because it delivers near best-in-class performance, handles quantile regression out-of-box, and it is computationally lightweight which means it can easily be developed as PoC without the need for high performance compute.

Explaining predictions from the ML model: finally, SHAPely additive explanations were used to deliver some insight into why the LightGBM model had arrived at its predictions. SHAPely additive explanations use Graph Theory to break down the model predictions into a baseline and a series of deviations from baseline caused by each feature in the model.

Results
The dynamic reserve setting model has been shown to outperform the existing approach to setting reserves because it is better able to flex reserve based on real-time data and short-term forecasts. Overall, the volume of reserve recommended is similar to the existing approach however in any given settlement period the reserve recommendations better follow NGESOS risk appetite.

Once the explainable ML model had been fully developed, the Smith Institute delivered a well-structured, documented, and production-ready suite of software for training and running the explainable probabilistic ML models for setting both upwards and downwards reserve levels. The approach adopted and the quality of the deliverables produced int the course of the project has ensured the models could be implemented into the business quickly and easily.

Project Review
The Dynamic Reserve Setting project can be considered a success for the following reasons.

- It demonstrated that it is possible to build an explainable probabilistic machine learning model which can be used to set operational

reserves and this model makes use of the latest real-time and forecast data so that the reserve recommendations are demonstrably better at meeting the company risk appetite when compared with the current approach for setting reserves.

- It delivered a software suite which once implemented internally will fully enable both real-time operation and periodic re-training of the ML models.

- It delivered an example ETL pipeline for transforming the data held within NGESO systems into a cleaned format stored in a MySQL database which can be used to support implementation within NGESO.

Project Dissemination

- https://www.current-news.co.uk/news/national-grid-eso-eyes-dynamic-day-ahead-reserve-setting-in-smith-institute-project
- https://www.nationalgrideso.com/news/national-grid-eso-and-smith-institute-begin-industry-pioneering-dynamic-reserve-setting-drs
- https://www.smithinst.co.uk/insights/national-grid-eso-and-smith-institute-begin-pioneering-drs-project/
- The Dynamic Reserve Setting project was presented at the 2021 Energy Networks Innovation Conference.
- All project reports are available on the Smarter Networks Portal.

Required Modifications to the Planned Approach During the Course of the Project

A dashboard to display the reserve recommendations and their SHAP explanations was not developed in the course of the project as previously planned. The data cleansing carried out early in the project required a significant amount of project time which affected the development of the dashboard. The SHAP values have been saved to html files, but no interactive application or dashboard has been constructed to explore the results.

Lessons Learnt for Future Projects

One lesson learnt for future projects is that the data cleaning and database curation took longer than anticipated. A significant amount of dialogue was required between NGESO and The Smith Institute to enable The Smith Institute to understand the internal NGESO data and be able to clean and store it appropriately. Although the data cleaning did not delay this project, future projects should take care to make sure they schedule an appropriate portion of project time towards data preparation.

A further lesson learnt is that SHAP values are a good format for providing explanations of predictions made by ML models to the control room engineers. Control room engineers cannot work with data from ML models which are "black box" and offer no justification for their predictions. However, stakeholder feedback from control room engineers about the use of SHAP values has been very positive and they should be considered for future innovation projects which seek to leverage ML.

Note: The following sections are only required for those projects which have been completed since 1st April 2013, or since the previous Project Progress information was reported.

The Outcomes of the Project

The Dynamic Reserve Setting has achieved the following:

- Built an ETL pipeline for cleaning data from NGESO systems to be used in reserves forecasting
- Built a database to store the cleaned data and forecast predictions
- Delivered a software suite which can run and re-train a probabilistic ML model for reserves levels
- Demonstrated that Machine Learning can be used to make better predictions of reserves when compared to the current approach
- Successfully used SHAP values to provide explanations for each prediction made by the Machine Learning model and these explanations were well received by the control room engineers

Data Access

Details on how network or consumption data arising in the course of a NIC or NIA funded project can be requested by interested parties, and the terms on which such data will be made available by National Grid can be found in our publicly available "Data sharing policy related to NIC/NIA projects" and www.nationalgrideso.com/innovation.

National Grid Electricity System Operator already publishes much of the data arising from our NIC/NIA projects at www.smarternetworks.org. You may wish to check this website before making an application under this policy, in case the data which you are seeking has already been published.

Foreground IPR

The project outputs will be a proof of concept for a dynamic reserve setting model. There will also be a performance comparison report comparing the benefits.