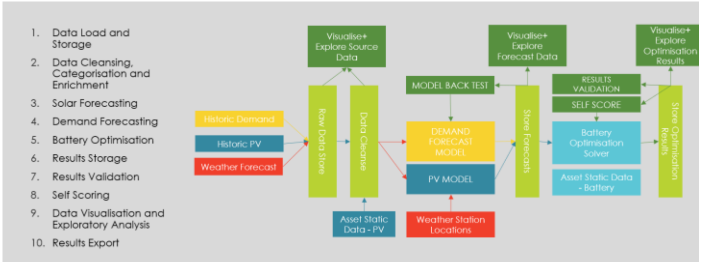
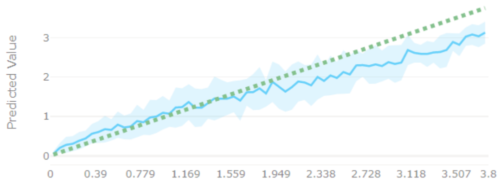
# Presumed Open Data – Method outline: Electric Dream Forecasters

## 1 Method Overview




**Calendar Map**: UTC timeslots were mapped to Winter / Summer, Local Time, Year, Month, Day of Week, Hour, Settlement Period, Bank Holidays and Working Days. This allowed the machine learning models to identify key time-based features.

**Training and Forecast Data Sets**: Task data was merged with the Calendar Map. Hourly weather forecasts were converted to half hourly by duplication of hourly values into each timeslot. A separate training data set was created for each task for demand and PV. The training data set included the Demand / PV values, the half hourly weather data and the calendar map *(NB: some correlation between demand and PV was noted, as would be expected but considered sufficiently reflected in the weather data to decouple the models)*. Forecast weather data was merged with the calendar map for the task week.




**Automated ML**: Azure Automated Machine Learning (AML) and the separate PV and Demand training data sets were used to test of the order of 50 different forecasting model setups each for PV and Demand per task. AML recommended the top models, and the individual model metrics were explored to assess accuracy and any tendency to under / over forecast. The preferred PV and Demand models were then deployed, providing an API to interact with the models and request forecasts based on time and weather values.

**Forecast Automation**: An Azure Function App was created to automate the creation of forecasts. This provided an API which received the task number and dates, from which the forecast weather data for the task week was automatically collated and input into the Demand and PV Forecast models and the results stored. This approach was introduced for later tasks and provided for quick comparison of different models and back testing for special days.

**Optimise by Hand**: An automated optimisation solver was considered; however, the task timelines (one optimisation per week) and nature of the optimisation were relatively simple. It was also considered that the success of the task would be based primarily on solar and demand forecast accuracy. Hence, a semi manual approach was adopted using Excel. At a high level the optimisation strategy attempted to reflect forecast uncertainty in solar and demand via various risk factors (these are described further under question 4). The optimiser also recreated the scoring function.

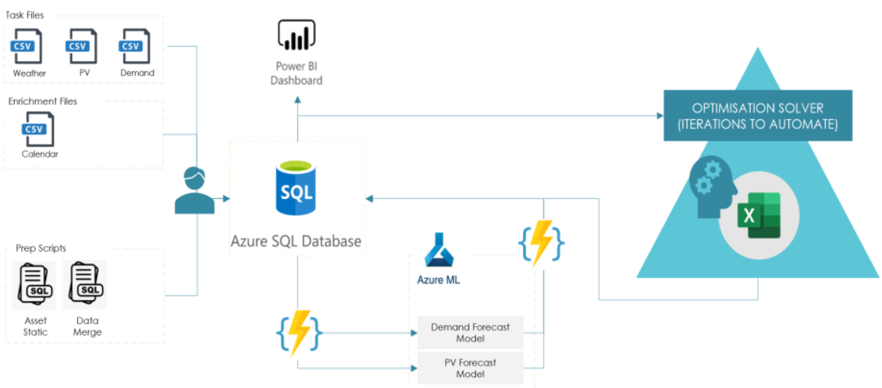
**Data Explorer:** Tooling for visualisation and analysis of source data, forecasts, and optimisation results was created.

**Review and Refine**: The team met once per week to review forecast and optimisation results and refine the approach prior to each task submission.

**Back Testing and Special Day Review**: Visualisation tools were utilised for review of task results (forecast vs actuals) and special day investigation.

## 2. Technical Design

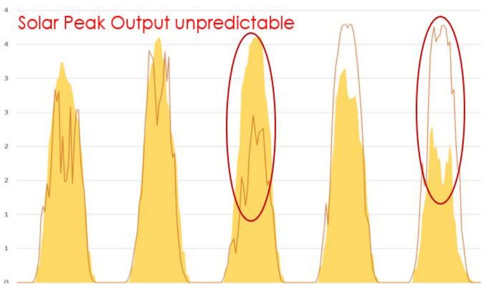
A bespoke solution was created for the purpose of the challenge. An SQL database for storage and manipulation of source data and results, Azure Automated Machine Learning for PV and Demand Forecasting model training and evaluation, Azure function app for automating data prep and forecast model interaction, a semi-automated optimisation solver in Excel, and Power BI for data visualisation and exploration.




## 3. Inputs

All task PV, Solar and Weather data was utilised for model training, except for cleaning of obvious erroneous values. PV Panel temperature and Irradiance were not used. Key features added were the local time and season mapping.

## 4. Iteration of Method





**Solar Strategy:** Forecast errors in solar output were identified as the key source of errors in the optimisation. Initial task dates were in the summer and most days were identified as 'high solar days' (i.e., solar power available during summer days exceeded that required to charge the battery). Risk factors were utilised to reflect the likely over forecasting of solar (due to unexpected cloud cover). This led to using only a subset of the forecast PV available for charging in any half hour. Later tasks with winter days contained predominantly 'low solar days' (i.e. solar power available was well below that sufficient to charge the battery). Here a charging envelope was created during day light hours to capture as much as possible of the available solar power. During task 1 a strategy of avoiding 'ramping hours' in the solar forecast and using more of the forecast PV in the peak hours was used – this was actually detrimental to results as it was found that the daylight hours are certain but the peak solar is highly uncertain (assumed due to difficulty to predict cloud cover).

**Correction for Demand Under Forecasting:** During task 2, it was identified that the demand forecast model was smoothing the peak and this was adversely impacting the optimisation results in terms of the peak reduction factor. A risk factor was introdued to the optimisation process to reflect demand under forecasting.

**Forecast Model Refinements:** Throughout the various tasks, PV and Demand Forecast models were re-trained, compared and assessed against actuals. Initially a Sparse Normalizer model was used for PV forecasting – this tended to under forecast at high values of PV output. Subsequently, Ensemble models were used. Ironically, a tendency to under forecast solar was often beneficial to the optimisation results (since over forecasting solar was the main reason for sub- optimal use of Grid Power, particularly on summer days).

## 5. Special Days – Christmas

No special adjustments were made to the demand forecast model / results for Christmas. Following review and analysis of historic data it was decided that forecasts from the models reflected a typical Christmas period.

## 6. Special Days – Covid

None

## 7. Data Use for Training and Testing

Following data cleansing, 100% of data was used for model training and back testing.

## 8. Comments / Suggestions

The challenge has been very well organised and enjoyable. We appreciate the efforts involved to co-ordinate the event and are impressed by the range of contestants from around the world. As energy industry experts but not professional data scientists, our aim was to have fun, try out some new technologies and meet people interested in the domain.

The challenge of manging peak distribution network demand is something our team is very aware of, working on projects with WPD such as Flexible Power, Intraflex and Sustain H. The challenge is somewhat simplified in terms of the optimisation problem and timelines typical in a real-world scenario. In practice, we optimise batteries within day - at least once every half an hour and drive PV / Demand forecasts based on live feedback from site (second by second real time data). The concept of a carbon intensity-based optimisation is interesting as we are typically price focused currently. It would be interesting to explore the correlation between price and carbon intensity (high carbon generation typically used in higher priced periods) – which could give a less binary grid / solar carbon weighting. Ultimately an efficient carbon scheme would appear the right mechanism to of ensuring price driven optimisation for carbon intensity.

As a team working full time and completing the challenge at evenings and weekends, time for refinement of the approach was limited, particularly between later tasks. With more time available we would focus on end-to-end automation of the process and the optimisation solver. Ideally a fully automated near real time forecast and optimisation solution would be achievable for real world application.

We would appreciate opportunity to interact with other teams after the challenged and discuss team's different approaches. In the spirit of open data, we have published our results and method online for all at:

https://tbowcutt.wixsite.com/podchallange