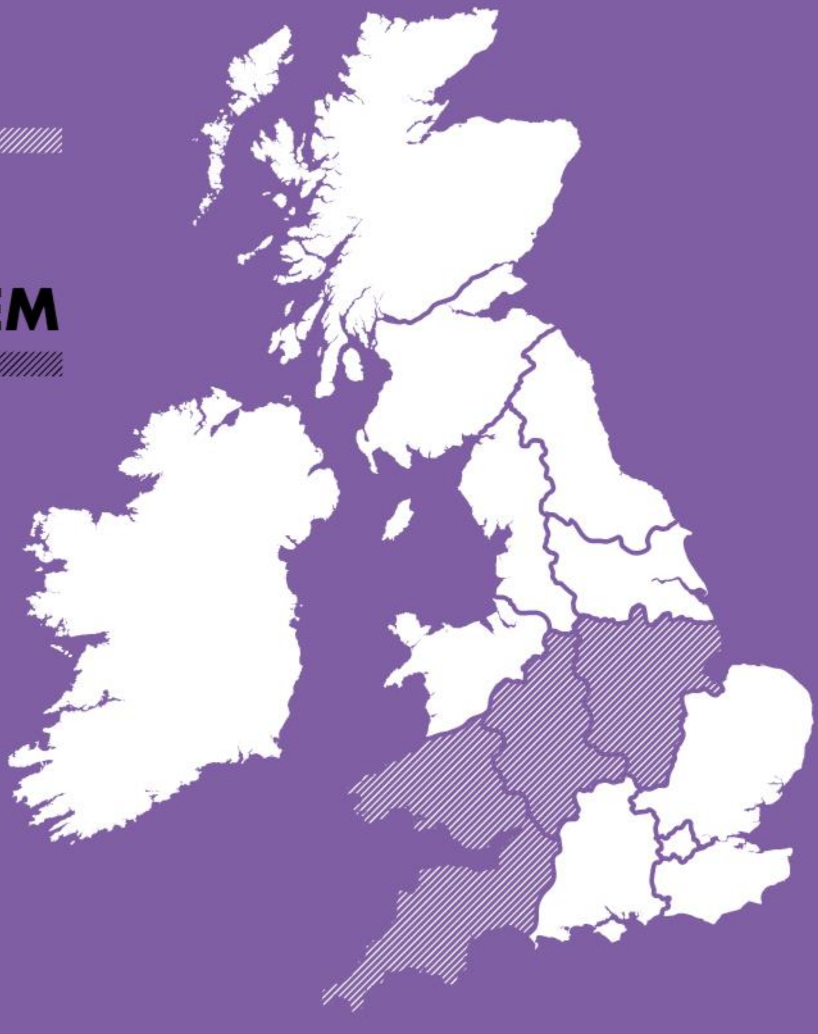
WESTERN POWER DISTRIBUTION

EFFS

ELECTRICITY FLEXIBILITY AND FORECASTING SYSTEM



EFFS
WPD/EN/NIC/03

NIC PROJECT FORECASTING VALIDATION TESTING REPORT

RIIO NIC NETWORK INNOVATION COMPETITION

EFFS

Report Title	:	EFFS Forecasting Validation Testing Report
Report Status	:	Final
Project Reference:	:	WPD/EN/NIC/03
Date	:	17 July 2019

Document Control		
	Name	Date
Prepared by:	Ivan Todorovic, Angela Diaz Rodriguez, George Marsh	17/07/2019
Reviewed by:	Jennifer Woodruff, Mike Pearson, John Hayling, David Pratt,	
Approved by:	Jennifer Woodruff	

Revision History		
Date	Issue	Status
04/07/2019	A	Initial draft
15/07/2019	B	Second draft
17/07/2019	C	Final

Contents

Contents 3
Glossary 5
Related Documents 6
1. Executive Summary 7
2. Introduction 8
2.1. EFFS Project Background 8
2.2. Objectives and Scope 8
2.3. Key Deliverables 11
3. Overseeing Model Development 12
3.1. Data Preparation 12
3.2. Model Development 12
3.3. Potential Improvements of the Forecasting Methodology 13
3.4. Overall Fitness for Purpose 16
4. Replication of Environment and Results 17
4.1. Technical Environment 17
4.2. Use Cases and Test Scenarios 17
4.3. Accuracy Metric and Acceptance Criteria 19
4.4. Comparison of Results 19
5. Results of Validation Testing 21
5.1. GSPs 21
5.2. BSPs 23
5.3. Primaries 26
5.4. Generation Customers 28
5.5. Load Customers 30
SCADA Data 30
Durabill Data 32
6. Comparison of Predicted vs. Actual Values 35
6.1. GSPs 36
6.2. BSPs 40
6.3. Primaries 48
6.4. Generation Customers 60
6.5. Load Customers 63
7. Conclusions and Recommendations 69
7.1 Replication of SGS toolchain and results 69
7.2 High-level conclusions 70
7.3 Recommendations for transition to BAU 71

DISCLAIMER

Neither WPD, nor any person acting on its behalf, makes any warranty, express or implied, with respect to the use of any information, method or process disclosed in this document or that such use may not infringe the rights of any third party or assumes any liabilities with respect to the use of, or for damage resulting in any way from the use of, any information, apparatus, method or process disclosed in the document.

Telephone +44 (0) 1332 827446. E-mail wpdinnovation@westernpower.co.uk

Glossary

Term	Definition
ARIMA	Auto Regressive Integrated Moving Average
BAU	Business as Usual
BSP	Bulk Supply Point
Capita DA	Capita Design Authority (Capita Employee Solutions Data Science Team)
DNO	Distribution Network Operator
DSO	Distribution System Operator
Durabill	WPD system that holds half-hourly metering data primarily for billing purposes
EFFS	Electricity Forecasting and Flexibility System
ENA	Energy Networks Association
GPU	Graphical Processor Unit
GSP	Grid Supply Point
I/O	Input / Output
LSTM	Long Short Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
NIC	Network Innovation Competition
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Networks
SCADA	Supervisory Control and Data Acquisition; a system that holds half-hourly metering data at multiple points in WPD's network
Stochastic	Having a random pattern that may not be predicted precisely
SGS	Smarter Grid Solutions
WPD	Western Power Distribution
XGBoost	Extreme Gradient Boosting

Related Documents

Ref	Document title	Version	Date issued	Prepared by	Location
1	EFFS Forecasting Report	Final	31/05/2019	SGS	https://www.westernpower.co.uk/projects/effs

1. Executive Summary

As the name suggests, the Electricity Flexibility and Forecasting Project, EFFS, involves work on forecasting. To purchase, arm and dispatch flexibility services effectively, Distribution System Operators (DSOs) will need to assess future power flows over their network and this requires forecasts for future load and generation values at different time horizons. Smarter Grid Solutions (SGS) provided an assessment of forecasting methods for EFFS. This report covers the validation testing carried out by Capita acting as the Design Authority (Capita DA) on the methods and tool chain proposed by SGS.

Capita DA has performed multiple simulations of forecasts using the SGS methodology on a range of locations in the WPD network, including GSPs, BSPs, primaries, load customers and generation customers and across the time horizons envisaged by the project (hour ahead to six months ahead). The results obtained were observed in terms of providing clarity on the following questions:

- Consistency of prediction accuracy over time for a given location and time horizon;
- Consistency of prediction accuracy for locations of the same type; and
- Comparison of prediction accuracy between types of location and time horizons.

Capita DA's key observations in this exercise are summarised below:

- The environment and models developed by SGS are based on open source technology and can be independently replicated to obtain equivalent results when applied on the same locations;
- Introducing a data cleansing process prior to producing forecasts is essential and Capita DA recommends that DSOs review their data quality and data collection as a first step. A specific tool for dealing with outliers and missing values is not part of the SGS toolchain and will need to be developed separately by DSOs;
- The model consistently met the defined acceptance criteria for a number of BSPs and primaries, even for the longer time horizons. It was observed that these locations had adequate data quality and a visible load pattern that could be replicated by the model;
- For other types of location, it was typically observed that the underlying behaviour is more random and the model could only meet acceptance criteria for the shortest time horizons;
- The quality of forecasts is highly dependent on the underlying behaviour of the location and Capita DA recommends that forecasts need to be optimised location by location (vs. applying general rules for the same type of location);
- Optimising a forecast for a specific location would first of all require adequate data quality, followed by observation of the underlying behaviour and experimentation with input features and model parameters (e.g. length of training data used). This functionality can be built into the toolchain and would be reasonable to perform given the speed of the XGBoost model.

In terms of transition of forecasting to a BAU processes, Capita DA believes the toolchain provides a starting point that can be adopted and applied by DNOs without the need for extensive specialised knowledge of forecasting. Further attention is needed to optimise the models for a specific location and time horizon. Finally, DSOs will need to establish a data I/O process to host the data and the forecasting models.

2. Introduction

2.1. EFFS Project Background

Western Power Distribution (WPD) is currently working on the Ofgem Network Innovation Competition (NIC) funded project Electricity Flexibility and Forecasting System (EFFS), with a projected end date of January 2021. This is a key project in their transition from Distribution Network Operator (DNO) to Distribution System Operator (DSO) and has the following objectives:

- Enhance the output of the Energy Networks Association (ENA) Open Networks project, looking at the high-level functions a DSO must perform, provide a detailed specification of the new functions validated by stakeholders, and the inclusion of specifications for data exchange;
- Determine the optimum technical implementation to support those new functions;
- Create and test the technical implementation by developing software and integrating hardware as required; and
- Use the testing of the technical implementation, which will involve modelling the impact of flexibility services to create learning relevant to forecasting, the likely benefits of flexibility services and the impact of changing network planning standards.

The EFFS project aims to design and implement a system which will allow the planning and dispatch of flexibility services in operational timescales. To do so, EFFS will use forecasts of generation and demand at specific network locations to drive the analysis of what those patterns mean for the distribution network.

As part of the EFFS project, WPD is seeking the development of a forecasting system, able to provide forecasts for load and generation at a range of timescales from an hour ahead to six months ahead, at various points in the DNO network. The intention of this project is to provide reliable, repeatable forecasting methods and algorithms to support the development of forecasting capacity. It is WPD’s intention that the learning and methods or algorithms will be transferable to the related NIC projects TRANSITION and FUSION, managed by Scottish and Southern Energy Network and Scottish Power Energy Networks respectively.

Smarter Grid Solutions (SGS) was selected as the forecasting partner in this project and has explored a number of forecasting methods. SGS has developed a toolchain that can be applied across all locations and time horizons using the XGBoost algorithm. Their methodology and results obtained are documented in the ‘EFFS Forecasting Report’ referred to in the Related Documents section. This report should be read in conjunction with the EFFS Forecasting Report.

2.2. Objectives and Scope

Capita DA’s scope of work consisted of two phases:

Phase 1 – overseeing model development. This work was performed during SGS’ development work and involved:

- Oversight of work performed by SGS in relation to EFFS project criteria and fitness for purpose of the models developed;

- Testing interim models as released by SGS and ensuring environment, models and results can be replicated; and
- Providing feedback to SGS on Capita DA's findings.

Phase 2 – validation testing. This phase of the project was performed once the final model and EFFS Forecasting Report were released by SGS and involved validation testing of the final SGS output. The aim of the validation testing work described in this report was to:

- Replicate the environment used by SGS and ensure the notebooks supplied by SGS run as intended;
- Replicate the results obtained by SGS on the same data and using the same models:
  - Capita DA intended to independently verify that models supplied by SGS, tuned and trained on the same data as used by SGS, yields equivalent prediction accuracy; and
  - This ensures that the model methodology has been successfully replicated.
- Apply the SGS models to a wider sample of locations in the WPD network, including:
  - GSPs;
  - BSPs;
  - Primaries to 33kV level;
  - Generation customers; and
  - Large load customers.
- Forecast the parameters above across for the following time horizons:
  - Hour ahead (i.e. the next two half-hourly readings);
  - Day ahead;
  - Week ahead;
  - One month ahead; and
  - Six months ahead.
- Apply the WPD-defined accuracy evaluation methods to calculate the efficacy of the forecasting methods;
- For each of the selected locations and each time horizon:
  - Tune the XGBoost model once;
  - Perform multiple forecasting simulations at different points in time within the data provided by WPD; and
  - At each simulation, train the model on past data, run predictions on unseen data and measure forecast accuracy against actual readings for the test period.
- Aggregate results over the simulations performed and observe:
  - The range of prediction accuracies over the simulations;
  - The variation in model performance across locations of the same family; and
  - The variation in model performance between location families and between time horizons.
- Based on the analysis above, described the results observed in relation to:
  - Conclusions drawn on how well the models perform on certain types of location and time horizon;
  - Insights into possible reasons behind variations in model performance;
  - Recommendations for improvement in model performance; and
  - Recommendations for implementation of forecasting by DNOs into BAU processes.

Aim of validation testing

Two key requirements of the EFFS forecasting model are:

1. Repeatability by DSOs; and
2. Based on open source tools.

With these in mind, Capita DA's first task was to replicate the environment used by SGS and ensure that models developed by SGS can be applied by Capita DA independently and yield the equivalent outcome. Secondly, Capita DA sought to verify that the models delivered by SGS (when trained and tested on the same data as SGS and using the same parameters) yield equivalent results.

A further requirement of the design authority work is to support WPD (as well as other DSOs and project stakeholders) in the adoption of the forecasting models in their BAU processes. To this end, it is important to gain an understanding of how the models perform on a broad sample of network locations and at different points in time.

Some of the answers that the validation testing exercise sought to provide are as follows:

- Consistency of prediction accuracy over time for a given location and time horizon;
- Consistency of prediction accuracy for locations of the same type (e.g. BSPs) across the network; and
- Comparison of prediction accuracy between types of location (e.g. BSPs vs. primaries) and time horizons (e.g. short-terms vs. long-term).

To support this analysis, Capita DA has developed a real-world forecasting simulation, applying the SGS model methodology on raw data provided by WPD. The model was applied to a broader sample of locations, including GSPs, BSPs, primaries, load customers and generation customers, across all five time horizons envisaged by the project.

Validation testing was performed on multiple time-splits for each selected location and time horizon combination in order to simulate the model's performance with the passage of time. The testing typically included six simulations for 6-month forecasts and 20 simulations for all other time horizons.

Results of the validation testing were assessed in terms of the overall reliability and robustness of forecasts and recommendations for adoption of forecasts into BAU processes by DNOs.

Scope of validation testing

The scope of Capita DA's validation testing exercise is described below:

- Consider a wider sample of locations including GSPs, BSPs, primaries, load customers, generation customers, across all five forecasting time horizons;
- Use the XGBoost toolchain to predict active power in MW (reactive power in MVar and other models out of scope due to time constraints);
- Tuning the XGBoost model once for a specific location and time horizon combination;
- Run multiple training and prediction sets at different points in time to simulate a real-world forecasting scenario, and report on observed results; and
- Follow the SGS methodology without applying additional processes (e.g. data preparation or introducing new features)

2.3. Key Deliverables

The key deliverables for the work performed by Capita DA were:

- This report, detailing the results of the evaluation testing; and
- Scripts showing the input of raw data, model tuning, model train/test simulations and results obtained.

It should be noted that Capita DA's work was focused on the application of the SGS models on WPD's raw data and observation of results obtained. Capita DA did not perform additional tasks that were not contained in the SGS toolchain, such as a data cleansing processes or performing multiple simulations on a single location in order to optimise results. The modifications performed by Capita DA were limited to:

- Data I/O process in order to pass the relevant time series as inputs to the SGS toolchain; and
- Creating a set of time splits for each time horizon, and a loop that allows the model to be re-trained and re-run for each time split.

3. Overseeing Model Development

3.1. Data Preparation

Capita DA was supplied with raw data from WPD covering a large number of network locations. The data typically contained readings in half-hourly intervals, for the years 2014-2018.

The raw data was split into training and test sets, with the training data provided to SGS to perform their modelling, and allow Capita DA to perform independent validation testing on the complete data set, part of which was unseen by SGS.

3.2. Model Development

Capita DA followed SGS' model development sprint by sprint. Observations with respect to the models developed are described below.

ARIMA – SGS spent several sprints developing and tuning the ARIMA model in R language, and attempted to reconstruct the model in Python. In some cases the ARIMA model provided adequate results, comparable to XGBoost. The ARIMA models were ruled out of final modelling, and some of the reasons for this were:

- Long training times for longer time horizons, making the model impractical for BAU processes; and
- Significant user skill and interaction required to tune the ARIMA model to each specific use case, resulting in a lower degree of automation compared to other models.

Capita DA has not been able to fully replicate the ARIMA model in R due to the specialised skills required in this specific domain. Capita DA's view was that ARIMA models would also be resource intensive for DSOs to replicate, adopt and maintain and therefore unsuitable from a fitness for purpose perspective.

XGBoost – this method was established by SGS to be faster and more flexible compared to other methods. Capita DA's validation testing confirmed that XGBoost could be readily adapted to different use cases and that tuning, training and prediction speeds were satisfactory (typically a few seconds to train and predict). Further, the forecasting accuracy was equal or better to the other models for each use case tested. Capita DA was satisfied with the fitness of purpose of XGBoost for the following reasons:

- Model can be automated for tuning and training/predictions;
- The level of user interaction is significantly lower compared to ARIMA and LSTM. Running and maintaining the model can be reasonably performed e.g. by an engineer with basic to intermediate Python and statistics skills;
- Tuning and training times allow for a large number of models to be performed on readily available hardware (e.g. all of Capita DA validation testing was performed on a standard laptop computer); and
- Model could be applied to any of the data sources explored, hence scalable across the DNO network.

LSTM – this method was explored by SGS and tested by Capita DA. The key concern was length of tuning and training time, rendering the model impractical for BAU. Although LSTM speed can be improved by running on GPU-equipped hardware, this was not explored further as the LSTM model did not offer tangible accuracy improvements over the XGBoost method.

3.3. Potential Improvements of the Forecasting Methodology

Data Quality

The SGS toolchain did not propose a data audit or data cleansing process, as raw data was used in model development and SGS base its work on a few selected use cases. In its validation testing, Capita DA has identified cases where data quality has clearly impacted forecast accuracy and the toolchain did not include processes which would address this.

Capita DA believes that data cleansing will be necessary in BAU forecasting, however this is a process that will be dependent on the specific data used by individual DNOs and developing robust methods for data cleansing are subject to further development and testing by DNOs. For this work package, the focus was on developing and testing a suitable forecasting method, rather than a complete end-to-end process.

Hyperparameter Tuning and Automation

The toolchain delivered by SGS included two Python scripts, each within a Jupyter notebook:

- One for tuning XGBoost hyperparameters for a specific use case
- One for training the XGBoost model and performing predictions on a specific use case

In performing validation testing, Capita DA has identified a number of downsides to this approach:

- In both notebooks the input data, features used, and dates need to be input separately. This increases the chances of human error and slows down the implementation of the model; and
- Once the tuning is performed the hyperparameters need to be entered manually into the training and forecasting notebook. This leaves room for improvement in the automation of the model – e.g. a single notebook with the option to perform tuning, with automated update of hyperparameters, would be more convenient for the user.

A further analysis of the sensitivity of model performance to hyperparameter tuning has not been explored in detail, hence the user sees hyperparameter optimisation as a 'black box' and is not provided with tangible insight into the importance or effect of parameters on the model. Capita DA believes that a default set of hyperparameters, e.g. for a location family or time horizon, would be a convenient feature for the user to have as a starting point – e.g. selecting a set of hyperparameters that worked well for a location of the same type.

The XGBoost implementation notebook user could then be easily amended to give the user the option of either using a default set of hyperparameters, or switching on the option to perform tuning and use the tuned hyperparameters instead.

Accuracy Metric

Further, a number of different accuracy metrics can be considered for measuring model performance. In this case the reported metric is the acceptance criteria (see section 4.3) based on the MAPE for consistency with the EFFS Forecasting Report. However during the validation testing other metrics were considered, including the MAE in MW readings, as well as the MAE expressed as a percentage of the nominal transformer capacity. The various metrics can all be easily implemented in the Python notebook, with the user deciding which option(s) to consider for the location in question.

Feature Selection

The SGS methodology is not fully prescriptive on what features constitute an optimal choice for a specific location and leaves room for the user to decide which features might be most useful in explaining the location's underlying behaviour. The methodology does provide some guidance – as described in the EFFS Forecasting Report and implemented again in this exercise (see Table 1):

- Temporal features and holidays only are used in the GSPs, BSPs, primaries and load customers;
- Wind data and temperature are added to the wind generation sites and some temporal features (e.g. day of the week) and holidays are removed.

In optimising the model for a specific location, Capita DA would recommend exploring the feature selection in more detail and in particular using the DSOs domain expertise in understanding the factors driving the behaviour of the location in question (e.g. the type of generation and load customers connected to the location).

Table 1. Default Feature Set by Type of Location

		Type of Location					
		GSP	BSP	Primary	Wind Farm	Solar Farm[1]	Large Load Customer
Features used as default	Hour	Yes	Yes	Yes	Yes	Yes	Yes
	Day of Week	Yes	Yes	Yes	No	No	Yes
	Quarter	Yes	Yes	Yes	Yes	Yes	Yes
	Month	Yes	Yes	Yes	Yes	Yes	Yes
	Year	Yes	Yes	Yes	Yes	Yes	Yes
	Day of Year	Yes	Yes	Yes	Yes	Yes	Yes
	Day of Month	Yes	Yes	Yes	Yes	Yes	Yes
	Week of Year	Yes	Yes	Yes	Yes	Yes	Yes
	Holidays	Yes	Yes	Yes	No	No	Yes
	Temperature	No	No	No	Yes	Yes	No
	Wind Output	No	No	No	Yes	Yes	No
	Wind Speed	No	No	No	Yes	Yes	No

In practice, the user could experiment by switching selected features on and off and start building knowledge about what features contribute to the model's accuracy. As an example, wind generation locations were tested using a number of temporal features that are not necessarily useful – in this case using fewer temporal features (e.g. month of year and time

[1] Solar farm not included in the validation testing exercise

of day only) in addition to the wind features, might help make the model more robust and less prone to overfit.

The implementation would be largely a manual process, where the feature set as set out in the EFFS Forecasting Report and this report for different types of location are used as a starting point, and the user would apply their judgement and domain knowledge in order to modify the feature set and attempt to improve results. Capita DA believes that it is necessary to perform more extensive testing in order to build sufficient results for concrete recommendations.

Training Period for Different Time Horizons

As with feature selection, the methodology has shown that the amount of past data used for training influences model performance. Varying this parameter will likely lead to an optimal outcome for a specific location. Running the model through a distinct set of (e.g. 2-3) training periods can again easily be implemented within the Python notebook as an additional feature.

Recommended Process for Optimising Forecasts for a Specific Location

With regards to optimising results for a given location, Capita DA's overall recommendation is to extend the SGS methodology to cover a wider range of overall parameters. As explained above, the XGBoost model hyperparameters are optimised, while the user needs to decide on the training period and features used.

Capita DA recommended approach would be to combine data cleansing, feature selection and model parametrisation in a process that would help yield optimal results for a specific location:

1. Examine data quality and check for outliers, extended periods of zero readings / error codes;
2. Apply a data cleansing process to the extent possible – Capita DA suggest a number of strategies:
    a. Occasional outliers can be interpolated from existing data;
    b. More extended periods of 'bad data' (e.g. several days at a time) can be manually avoided in the training data by reducing the training period to only include 'good data'. This may mean using the data for shorter-term forecasts only;
    c. Systematic periods of 'bad data' may require the data collection process to be amended first before this location can be used for forecasting.
3. Once the data is cleansed (if possible), visually inspect the behaviour to determine the randomness of its behaviour. The user can expect well-behaved locations to work reasonably well with default parameters, while more stochastic behaviour will likely require additional optimisation work;
4. Run a first set of simulations to observe results:
    a. Set up the XGBoost model using default training length and feature set;
    b. Apply tuning to obtain hyperparameters;
    c. Select the most appropriate error metric to follow (user's discretion);
    d. Run the model on several simulations through the data (e.g. 6-20 simulations as described in this report) and observe the range of results obtained on the selected error metric;

5. Observe the results and decide whether acceptance criteria are met (model is good enough), or model needs to be optimised further;
6. If optimisation is needed:
    a. Select a range of training lengths to try, and a set of features to be turned on or off. For e.g. three different training periods and three features to test, this would result in nine distinct combinations;
    b. For each of the combinations above, run the model through simulations and observe results;
    c. Select the model where the acceptance criteria are best met (the criteria for best model selection will depend on the user's preferences and could be e.g. the model where the acceptance criteria are passed the most, or where the variance between simulations is the least;
7. Decide if the accuracy of the best-performing model meets the user's acceptance criteria to be used in a BAU process.

The above procedure is exhaustive and would help ensure that the model selected has been optimised for features, training length and hyperparameters, and that it has been tested on multiple simulations across the available body of data.

Capita DA believes that building this functionality in the toolchain will require a limited amount of additional work, while providing a systematic way of optimising the model for a specific location. Applied one location at a time, it would lead to a robust set of location-specific models and a known degree of accuracy for each one.

Capita DA recommends that building models one location at a time is a prudent approach for DNOs. For application across a large number of locations, efficiencies should be sought in automating the above process, which could be performed by e.g. by an engineer with Python programming skills.

3.4. Overall Fitness for Purpose

Although the toolchain can be improved further, Capita DA considers it to be fit for purpose in relation to objectives for the EFFS forecasting work:

- The toolchain is based on open source technology and is replicable by DNOs;
- The model can be tuned and applied to any location in the network and provide forecasts for all of the time horizons considered (subject to availability of past data);
- The model tuning, training and prediction times allow forecasting to be applied on a large scale using readily available hardware;
- Capita DA has been able to apply the toolchain to a wide sample of locations;
- Cases where the model may perform better or worse have been observed and documented; and
- Capita DA considers that DNOs will be able to integrate the toolchain into a data I/O process and use the toolchain to establish a continuous forecasting activity.

4. Replication of Environment and Results

4.1. Technical Environment

The toolchain developed by SGS runs in an Anaconda environment. This consists of a distribution of the Python language and libraries commonly used in data science, machine learning, numerical computation and visualisation.

Capita DA has been able to replicate this environment, as well as additional libraries used in the toolchain by following instructions as described in the EFFS Forecasting Report. As with similar open source tools, it is advisable to check versions of packages installed in order to ensure the toolchain operates as intended.

SGS has also described in its report the installation procedure and use of a PostgreSQL database for input and output of the forecasting data, including the TimescaleDB extension to support time series data.

Capita DA did not test or use this interface for its validation testing procedure and therefore does not provide any assessment of the PostgreSQL and TimescaleDB setup in this report. The choice of this setup for the data input and output interface is left to individual DSOs to make based on their own preferences. The PostgreSQL solution is an open source tool that together with the TimescaleDB extension supports time series data as needed for the forecasting. However, a variety of common relational database solutions can be connected to Python and therefore DSOs may decide to implement a different solution. In that case, changes to the Jupyter notebooks would be required in the data I/O section, and typically a library would need to be installed and imported to interface between the database and Python.

4.2. Use Cases and Test Scenarios

In the EFFS Forecasting Report, SGS has reported prediction accuracies in relation to acceptance criteria for a total of seven use cases. Capita DA has performed its validation testing on six of the seven use cases and compared its results to those reported by SGS.

Table 2. Summary of Use Cases described in the EFFS Forecasting Report and tested by Capita DA

Use Case	Location	Time Horizons	Features	Data Inputs & Sources
UC1	Indian Queens GSP 4x 240 MVA Transformers Forecasts are produced for each transformer, and an aggregate produced by summing individual transformers.	Six Months Ahead Month Ahead Week Ahead Day Ahead Hour Ahead	Hour Day of Week Quarter Month Year Day of Year Day of Month Week of Year Holidays	Indian Queens SGP 180 – MW Indian Queens SGP 380 – MW Indian Queens SGP 480 – MW Indian Queens SGP 280 – MW Bank holidays for England and Wales[2]

[2] http://www.calendarpedia.co.uk

Use Case	Location	Time Horizons	Features	Data Inputs & Sources
UC2	Cardiff South BSP 2x 40 MVA Transformers Forecasts are produced for the aggregate BSP.	Six Months Ahead Month Ahead Week Ahead Day Ahead Hour Ahead	Hour Day of Week Quarter Month Year Day of Year Day of Month Week of Year Holidays	Cardiff SouthGRID 1Power MW Cardiff SouthGRID 2Power MW Bank holidays for England and Wales[1]
UC3	Prince Rock primary 2x 17.25 MVA Transformers Forecasts are produced for the aggregate primary.	Six Months Ahead Month Ahead Week Ahead Day Ahead Hour Ahead	Hour Day of Week Quarter Month Year Day of Year Day of Month Week of Year Holidays	PRINCE ROCKCB 27/19Power MW PRINCE ROCKCB 27/21Power MW Bank holidays for England and Wales[1]
UC4	Truro BSP 2x 60 MVA Transformers Forecasts are produced for the aggregate BSP.	Six Months Ahead Month Ahead Week Ahead Day Ahead Hour Ahead	Hour Day of Week Quarter Month Year Day of Year Day of Month Week of Year Holidays	TRURO BSPCB 1T0Power MW TRURO BSPCB 2T0Power MW (inverted as measurement appears to be in the wrong direction) Bank holidays for England and Wales[1]
UC5	Llynfi Valley primary 1x 12 MVA 1x 21 MVA Transformers Forecasts are produced for the aggregate primary.	Six Months Ahead Month Ahead Week Ahead Day Ahead Hour Ahead	Hour Day of Week Quarter Month Year Day of Year Day of Month Week of Year Holidays	LlynfiTrans 1Power MW LlynfiTrans 2Power MW Bank holidays for England and Wales[1]
UC6	Goonhilly Wind Farm, the Lizard, Cornwall 12 MVA Capacity	Six Months Ahead Month Ahead Week Ahead Day Ahead Hour Ahead	Hour Quarter Month Year Day of Year Day of Month Week of Year Temperature Wind Output Wind Speed	Goonhilly MW Temperature[3] Wind Output[2] Wind Speed[2]

[3] Renewables Ninja - https://www.renewables.ninja/ - for The Lizard, Cornwall.

4.3. Accuracy Metric and Acceptance Criteria

Prediction accuracy is calculated as the MAPE, explained by the following formula, as specified by WPD and used by SGS:

$$Accuracy\ (\%) = 100 - \left(\left|\frac{Actual - Prediction}{Actual}\right| \times 100\right)$$

The results reported show the percentage of predictions that pass the given accuracy threshold:

- *'> 50% accuracy threshold'* refers to the percentage of predictions in the prediction time horizon that fall between 50% and 150% of the actual value (i.e. the pass rate for the threshold)
- *'> 80% accuracy threshold'* refers to the percentage of predictions in the prediction time horizon that fall between 80% and 120% of the actual value (i.e. the pass rate for the threshold)
- The reported figures are the pass rates averaged over the simulations performed by SGS and Capita DA
- *'Difference'* shows the percentage point difference in the reported figures between Capita DA and SGS, i.e. how closely matched they are

The acceptance criteria are considered to be met when:

- The 50% accuracy threshold is passed more than 80% of the time; and
- The 80% accuracy threshold is passed more than 80% of the time.

4.4. Comparison of Results

Capita DA has applied the SGS methodology to the same use cases, and observed results over the simulations. For each use case and time horizon combination, the XGBoost model is tuned once, and simulations are run multiple times at different points in time. At each simulation the model retrains on a specified period of past data and runs predictions for the specified time horizon.

The table below shows a comparison of results achieved by SGS vs. those achieved by Capita DA:

Table 3. Comparison of accuracy reported by SGS and by Capita DA's validation testing on the same Use Cases

	>50% accuracy threshold			>80% accuracy threshold		
UC1 - GSP Indian Queens	**Reported by SGS**	**Validation Testing**	**Difference**	**Reported by SGS**	**Validation Testing**	**Difference**
6 months	30.6	15.9	-14.8	11.9	6.2	-5.8
1 month	28.9	23.8	-5.1	11.7	8.4	-3.3
1 week	25.1	33.6	8.5	9.4	14.6	5.1
1 day	31.0	25.6	-5.3	13.4	11.4	-2.0
1 hour	50.0	72.5	22.5	25.0	42.5	17.5
UC2 - BSP Cardiff South	**Reported by SGS**	**Validation Testing**	**Difference**	**Reported by SGS**	**Validation Testing**	**Difference**
6 months	99.4	99.2	-0.2	79.2	87.1	7.9
1 month	99.9	98.8	-1.2	83.5	84.6	1.1
1 week	99.8	99.5	-0.2	92.1	91.1	-1.0

1 day	100.0	98.3	-1.7	97.3	96.4	-1.0
1 hour	100.0	100.0	0.0	100.0	100.0	0.0
UC3 - Prince Rock primary	**Reported by SGS**	**Validation Testing**	**Difference**	**Reported by SGS**	**Validation Testing**	**Difference**
6 months	98.2	99.9	1.7	96.05	95.4	-0.6
1 month	100.0	99.7	-0.2	98.59	94.1	-4.5
1 week	100.0	100.0	0.0	99.33	98.6	-0.8
1 day	100.0	100.0	0.0	99.7	99.6	-0.1
1 hour	100.0	100.0	0.0	100	100.0	0.0
UC4 - BSP Truro	**Reported by SGS**	**Validation Testing**	**Difference**	**Reported by SGS**	**Validation Testing**	**Difference**
6 months	69.0	48.9	-20.1	29.9	19.9	-10.0
1 month	73.5	52.3	-21.2	33.8	23.4	-10.4
1 week	73.4	64.1	-9.3	34.1	32.0	-2.1
1 day	85.1	74.7	-10.4	45.5	44.1	-1.5
1 hour	100.0	100.0	0.0	52.1	95.0	42.9
UC5 - Llynfi primary	**Reported by SGS**	**Validation Testing**	**Difference**	**Reported by SGS**	**Validation Testing**	**Difference**
6 months	97.5	92.8	-4.7	87.4	67.7	-19.7
1 month	97.7	93.9	-3.8	87.0	72.5	-14.5
1 week	99.0	98.0	-1.0	91.4	93.3	1.9
1 day	100.0	98.0	-2.0	98.5	93.3	-5.2
1 hour	100.0	100.0	0.0	100.0	100.0	0.0
UC6 - Goonhilly Wind Farm	**Reported by SGS**	**Validation Testing**	**Difference**	**Reported by SGS**	**Validation Testing**	**Difference**
6 months	37.3	45.5	8.2	12.8	20.2	7.4
1 month	40.4	56.2	15.9	18.7	29.6	10.9
1 week	48.9	40.6	-8.3	27.5	15.3	-12.2
1 day	87.2	72.4	-14.8	71.7	36.7	-35.0
1 hour	87.5	86.8	-0.7	79.2	52.6	-26.5

The comparison of results reveals that, for use cases where prediction accuracy is high (e.g. UC2), the results achieved by SGS and Capita DA are very close:

- < 2 percentage points for time horizons of one month and below
- 7.9 percentage points for six months ahead

In use cases where prediction accuracy is lower, the variation in results between SGS testing and validation testing is greater – e.g. for UC4 hour ahead, 42.9 percentage point difference is observed. This is explained by the less predictable behaviour of this particular use case, which results in less reliable predictions.

5. Results of Validation Testing

Capita DA extended its validation testing beyond the use cases covered by SGS, in order to demonstrate model performance over a wider sample of locations. The results of this exercise are summarised in this section for each type of location (i.e. GSPs, BSPs, primaries, generation customers and load customers).

5.1. GSPs

Table 4. Testing Parameters for GSPs

Location	Time Horizons	Features	Tuning, Validation, Forecasting Periods	Data Inputs & Sources
GSP: **Indian Queens** **Landulph** Forecasts are produced for each transformer.	Six Months Ahead Month Ahead Week Ahead Day Ahead Hour Ahead	Hour Day of Week Quarter Month Year Day of Year Day of Month Week of Year Holidays	Six Months Ahead: 14-12-2015 to 06-08-2018 6 simulations Month Ahead: 30-06-2014 to 13-09-2018 20 simulations Week Ahead: 30-06-2014 to 20-08-2018 20 simulations Day Ahead: 01-06-2014 to 19-06-2018 20 simulations Hour Ahead: 01-06-2014 to 18-06-2018	Indian Queens SGP 180 – MW Indian Queens SGP 280 – MW Indian Queens SGP 380 – MW Indian Queens SGP 480 – MW Landulph Supergrid 180 – MW Landulph Supergrid 280 – MW Landulph Supergrid 380 - MW Bank holidays for England and Wales

Table 5. Simulation Parameters for GSPs

Time horizon	Tuning period	Validation period	Training period	No. of simulations
Six Months Ahead	11 months	1 month	12 months	6
Month Ahead	11 months	1 month	12 months	20
Week Ahead	11 months	1 month	12 months	20
Day Ahead	12 months and 3 weeks	1 week	13 months	20
Hour Ahead	12 months and 3 weeks	1 week	13 months	20

The GSPs tested include GSP Indian Queens (UC1) and GSP Landulph. In each case, each transformer within the GSP is modelled separately.

It can be observed that:

- The hour ahead forecasts reach the 80% accuracy threshold on average by six out of the seven transformers tested;
- The 50% accuracy threshold has been achieved on average for day ahead forecasts by four out of the seven transformers;
- The variation in results was significant in nearly all cases, hence even the hour ahead predictions need to be treated with caution; and

- The lack of prediction accuracy is explained by the high level of aggregation at GSP level, meaning that any unpredictable behaviour further down the network will propagate through to the GSP

A closer inspection of the predicted and actual values (section 6.1) reveals some additional insight:

- Data quality issues are present (e.g. extended zero readings in GSP Landulph TX2);
- The daily pattern of a transformer can vary significantly and the XGBoost model does not fully predict these variations (e.g. GSP Indian Queens – day ahead); and
- The actual readings do not vary around a steady mean, but can trend up or down, and the XGBoost model does not fully incorporate these trends (e.g. GSP Landulph TX2 – 1 month, Median Case – see Figure 14, Section 6.1).

Figure 1. Percentage of predictions passing the 80% accuracy threshold

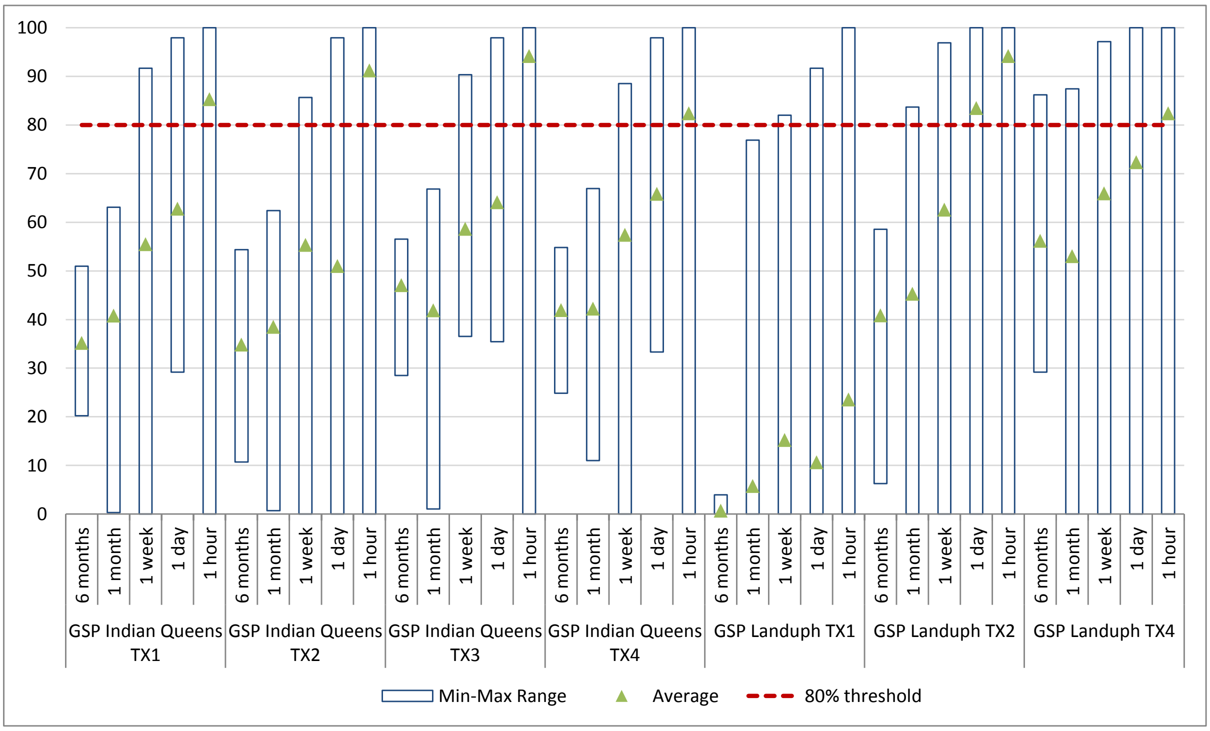
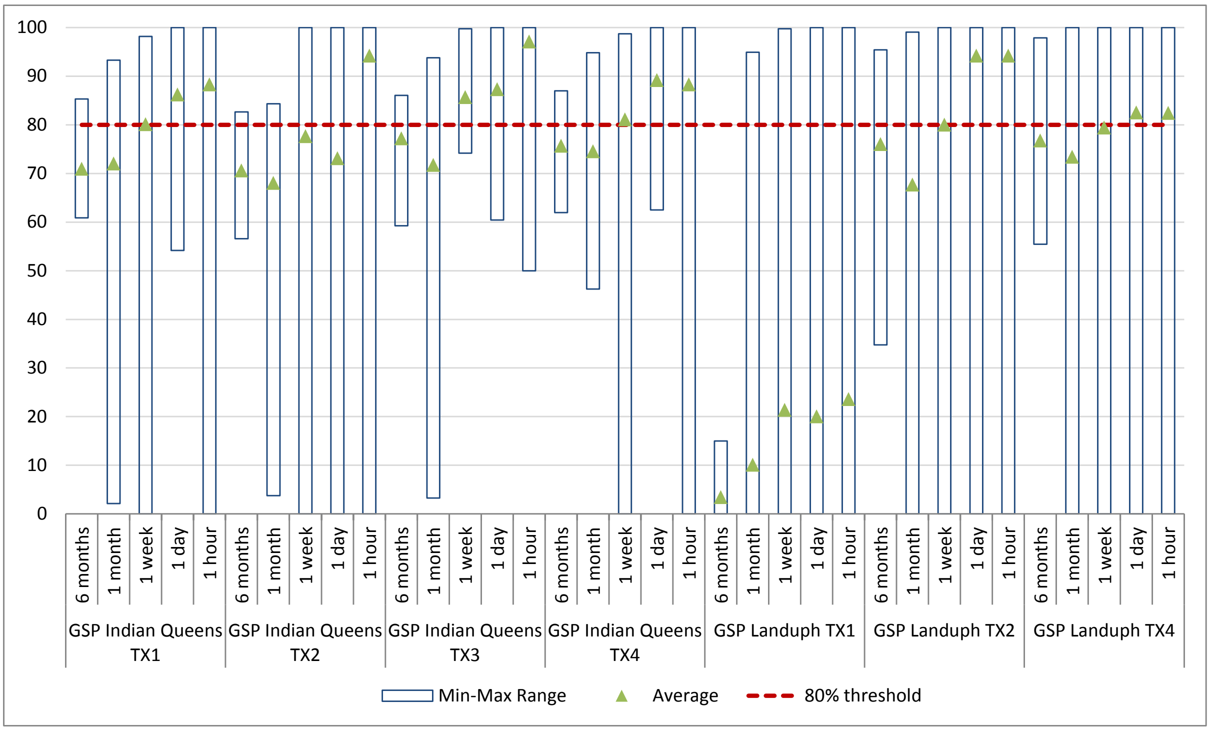


Figure 2. Percentage of predictions passing the 50% accuracy threshold



5.2. BSPs

Table 6. Testing Parameters for BSPs

Location	Time Horizons	Features	Tuning, Validation, Forecasting Periods	Data Inputs & Sources
BSP: **Cardiff South** **Morriston** **Truro** **Ludlow** Forecasts are produced for the aggregate BSP.	Six Months Ahead Month Ahead Week Ahead Day Ahead Hour Ahead	Hour Day of Week Quarter Month Year Day of Year Day of Month Week of Year Holidays	Six Months Ahead: 14-12-2015 to 06-08-2018 6 simulations Month Ahead: 30-06-2014 to 13-09-2018 20 simulations Week Ahead: 30-06-2014 to 20-08-2018 20 simulations Day Ahead: 01-06-2014 to 19-06-2018 20 simulations Hour Ahead: 01-06-2014 to 18-06-2018	Cardiff SouthGRID 1Power MW Cardiff SouthGRID 2Power MW MorritsonGRID 1Power MW MorristonGRID 2Power MW TRURO BSPCB 1T0Power MW TRURO BSPCB 2T0Power MW Ludlow 33kVGT2CPower MW Ludlow 33kVGT3Power MW Bank holidays for England and Wales

Table 7. Simulation Parameters for BSPs

Time horizon	Tuning period	Validation period	Training period	No. of simulations
Six Months Ahead	11 months	1 month	12 months	6
Month Ahead	11 months	1 month	12 months	20
Week Ahead	11 months	1 month	12 months	20
Day Ahead	12 months and 3 weeks	1 week	13 months	20
Hour Ahead	12 months and 3 weeks	1 week	13 months	20

Capita DA tested the XGBoost model on four BSPs – BSP Cardiff South (UC2), BSP Ludlow 33kV, BSP Morriston and BSP Truro (UC3). The model performed better on BSPs than on GSPs, as the behaviour at BSP level show more predictable patterns that allow the XGBoost model to yield more robust forecasts.

From results of the validation testing, it can be observed that:

- The 80% accuracy threshold is reached on average by all of the four BSPs, for the hour ahead time horizon;
- The 80% accuracy threshold is reached on average by three out of the four BSPs, for week ahead and day ahead time horizons;
- The 50% accuracy threshold is reached on average by three out of the four BSPs, for all time horizons;
- Prediction accuracies generally improve with shorter time horizons.

A closer inspection of the predicted and actual values reveals further insight:

- As an example, BSP Cardiff South - month ahead shows a predictable weekly profile that is closely predicted by the XGBoost model in the best and median cases; however a change in pattern is observed over the Christmas / New Year period (worst case);
- BSP Truro exhibits more variation in intraday readings compared to the other BSPs, leading to lower accuracy. In this case understanding the underlying reasons for this behaviour would be helpful in determining a suitable course of action for improving predictions;
- A poor set of predictions can be caused by error readings in the training data, e.g. BSP Morriston - month ahead – Worst Case Split 18 (see Figure 19, Section 6.2). The data was affected by negative readings (likely to be error codes) in the training data, forcing the model to learn wrong values. A data quality check is needed to flag cases such as this one prior to predictions being performed.

In the cases where aggregate-level model performance is inadequate, it may be worth experimenting with forecasts at transformer level, keeping in mind that these may be easier to adapt for non-standard network configurations.

Figure 3. Percentage of predictions passing the 80% accuracy threshold

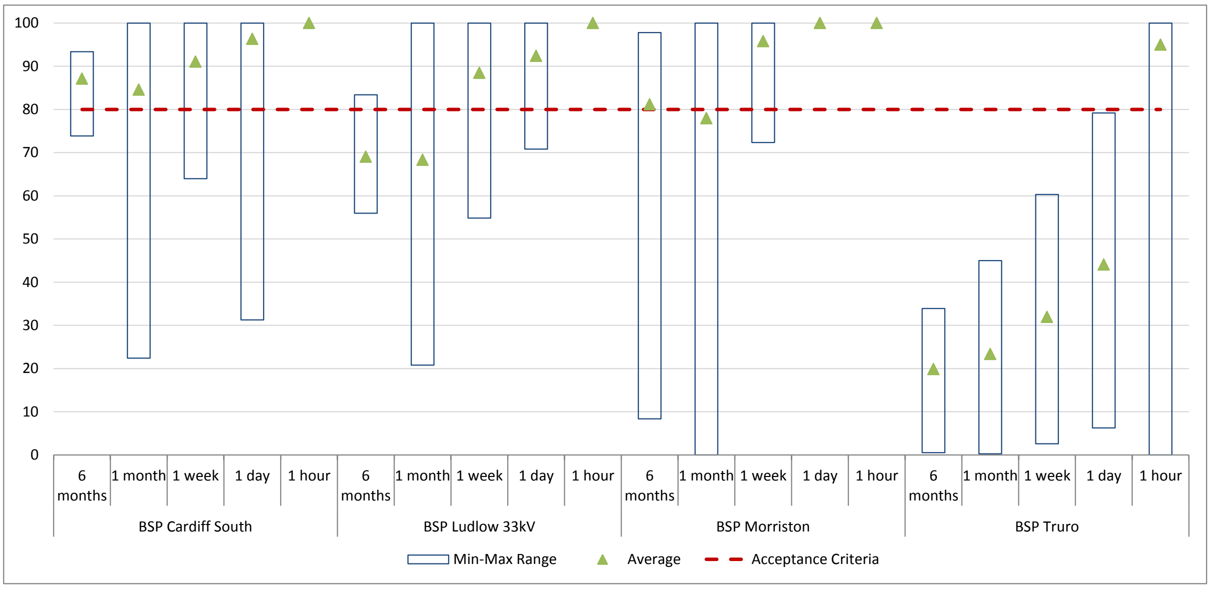
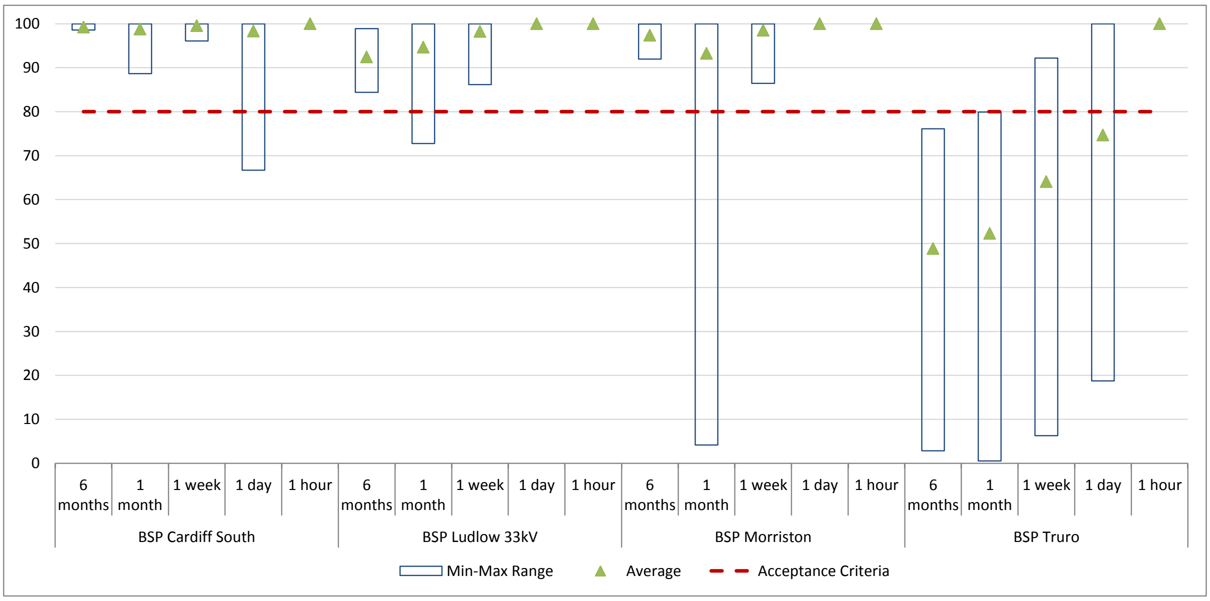


Figure 4. Percentage of predictions passing the 50% accuracy threshold



5.3. Primaries

Table 8. Testing Parameters for Primaries

Location	Time Horizons	Features	Tuning, Validation, Forecasting Periods	Data Inputs & Sources
Primary: **Prince Rock** **Kingsweston** **Evercreech** **Cardiff East** **Llynfi** **St Clears** Forecasts are produced for the aggregate primary.	Six Months Ahead Month Ahead Week Ahead Day Ahead Hour Ahead	Hour Day of Week Quarter Month Year Day of Year Day of Month Week of Year Holidays	Six Months Ahead: 14-12-2015 to 06-08-2018 6 simulations Month Ahead: 30-06-2014 to 13-09-2018 20 simulations Week Ahead: 30-06-2014 to 20-08-2018 20 simulations Day Ahead: 01-06-2014 to 19-06-2018 20 simulations Hour Ahead: 01-06-2014 to 18-06-2018	PRINCE ROCKCB 27/19Power MW PRINCE ROCKCB 27/21Power MW KINGSWESTONCB 6Power MW KINGSWESTONCB 8Power MW EVERCREECHCB 3Power MW EVERCREECHCB 5Power MW Cardiff EastGRID 1Power MW Cardiff EastGRID 3Power MW LlynfiTrans 1Power MW LlynfiTrans 2Power MW St ClearsTRANS 1Power MW St ClearsTRANS 2Power MW Bank holidays for England and Wales

Table 9. Simulation Parameters for Primaries

Time horizon	Tuning period	Validation period	Training period	No. of simulations
Six Months Ahead	11 months	1 month	12 months	6
Month Ahead	11 months	1 month	12 months	20
Week Ahead	11 months	1 month	12 months	20
Day Ahead	12 months and 3 weeks	1 week	13 months	20
Hour Ahead	12 months and 3 weeks	1 week	13 months	20

Validation testing was performed on six primaries: Evercreech, Kingsweston, Llynfi (UC5), Prince Rock (UC3), Cardiff East and St Clears. Overall the XGBoost model performed best on some of the primaries. The following observations can be made:

- For three primaries (Evercreech, Prince Rock and Cardiff East), the results exceeded the 50% accuracy benchmark for all time horizons and in all simulations. The 80% accuracy benchmark was reached for all time horizons on average, though some simulations fell short of the acceptance criteria;
- St Clears is affected by error readings on one of the transformers throughout the observed period, highlighting the need for a data check;
- Kingsweston and Llynfi exhibit less predictable behaviour, hence results trail those of the top three primaries. These are cases where a closer investigation of the factors driving underlying behaviour would be a natural next step.

A closer inspection of the predicted and actual values reveals further insight:

- In the best modelled primaries (Evercreech, Prince Rock and Cardiff East), the model is able to predict the general pattern and errors tend to occur in the daily peaks;

- For the two more difficult cases (Kingsweston and Llynfi), it the patterns are far less clear and there are periods of systematic prediction errors.

In the more difficult cases, it may be worth experimenting with transformer-level forecasts.

Figure 5. Percentage of predictions passing the 80% accuracy threshold

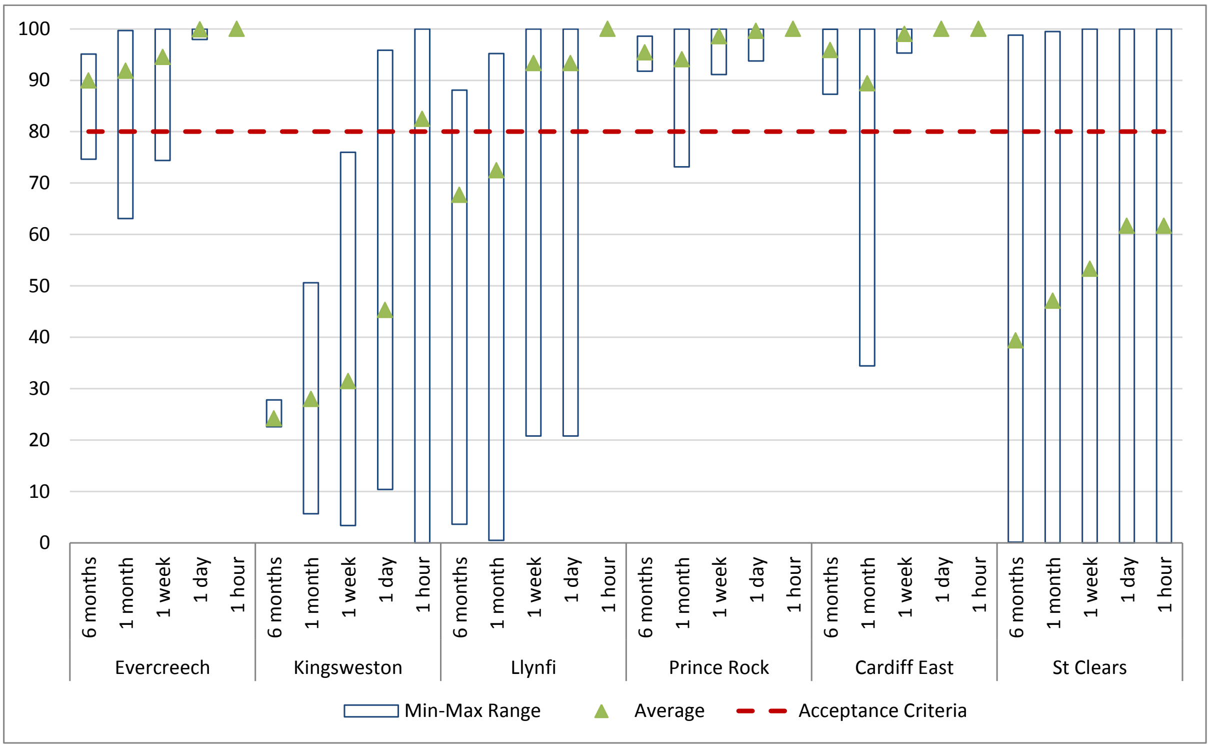
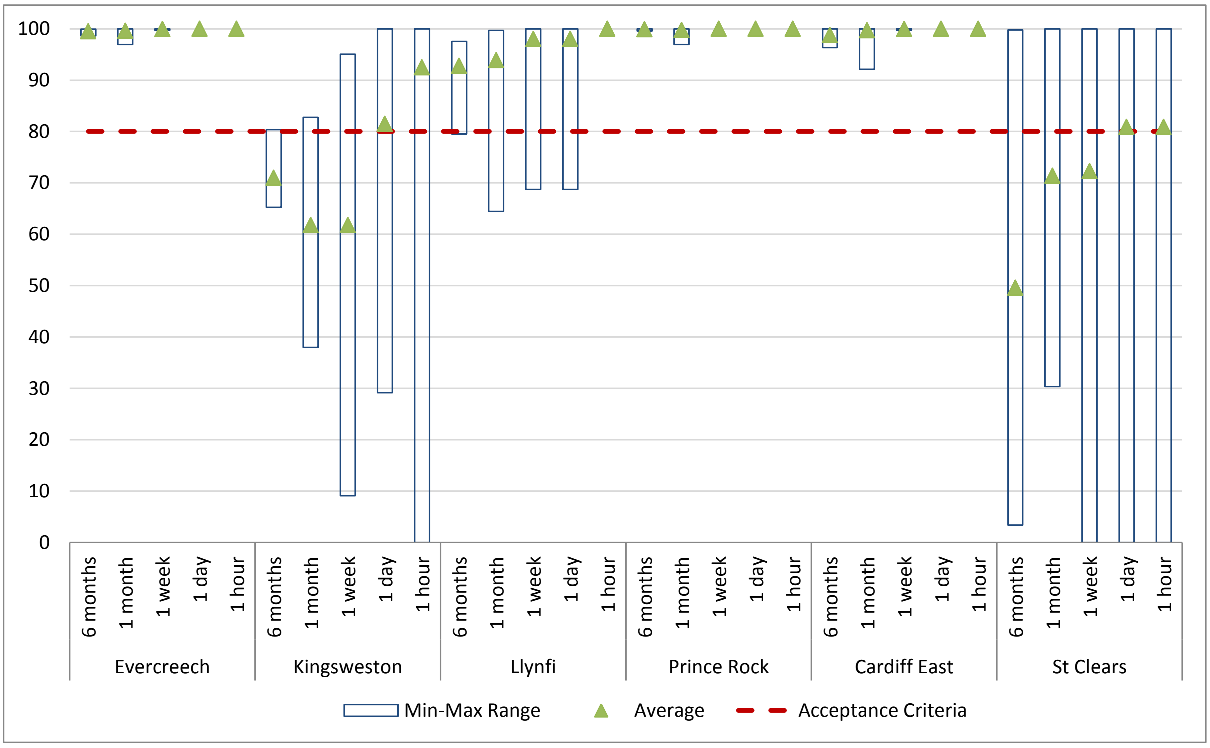


Figure 6. Percentage of predictions passing the 50% accuracy threshold



5.4. Generation Customers

Table 10. Testing Parameters for Generation Customers

Location	Time Horizons	Features	Tuning, Validation, Forecasting Periods	Data Inputs & Sources
Wind Farm: Goonhilly Rockhead	Six Months Ahead Month Ahead Week Ahead Day Ahead Hour Ahead	Hour Quarter Month Year Day of Year Day of Month Week of Year Temperature Wind Output Wind Speed	Six Months Ahead: 14-12-2015 to 07-09-2017 6 simulations Month Ahead: 14-12-2014 to 11-12-2018 19 simulations Week Ahead: 14-12-2014 to 05-12-2018 19 simulations Day Ahead: 14-11-2014 to 06-10-2018 19 simulations Hour Ahead: 14-11-2014 to 06-10-2018 19 simulations	Goonhilly MW Rockhead MW Temperature Wind Output Wind Speed

Table 11. Simulation Parameters for Generation Customers

Time horizon	Tuning period	Validation period	Training period	No. of simulations
Six Months Ahead	11 months	1 month	12 months	6
Month Ahead	11 months	1 month	12 months	19
Week Ahead	11 months	1 month	12 months	19
Day Ahead	12 months and 3 weeks	1 week	13 months	19
Hour Ahead	12 months and 3 weeks	1 week	13 months	19

Validation testing was performed on two wind farms: Goonhilly (UC6) and Rockhead. For these locations, wind speed and wind direction were added as features, based on data from the website renewables.ninja relevant to each wind farm site. It should be noted that his data refers to actual wind data rather than forecasts available at the time of making the predictions (therefore looking into the future). This is likely to result in higher levels of accuracy in the predicted wind output than if forecast data were used as weather forecast error is likely to be a significant factor, especially for the longer time-horizons. DSOs may wish to explore introducing forecast weather data in the features, if weather forecast data is available to obtain from a supplier. Due to time constraints in this project, only actual weather data was used.

Results of the testing show that only one of the wind farms could yield average forecasts above the 50% accuracy threshold. On closer inspection of the actual and predicted data, it is clear the behaviour is unpredictable and the model is struggling to identify any reliable pattern. SGS recommended the use of engineering models, such as those available from the renewables.ninja site, rather than creating time-series based forecasts using XGBoost for predicting the output of renewable generation because of the known non-linear features of this type of generation.

One idea for improvement may be to investigate the engineering models and the data available to DNOs to select the most appropriate version of these models e.g. manufacturer and type of turbine, height of the nacelle above ground level etc. If time-series forecasting is used, a further alternative might be to investigate whether the half hourly metering data for generators provides better accuracy than the SCADA monitoring data.

Figure 7. Percentage of predictions passing the 80% accuracy threshold

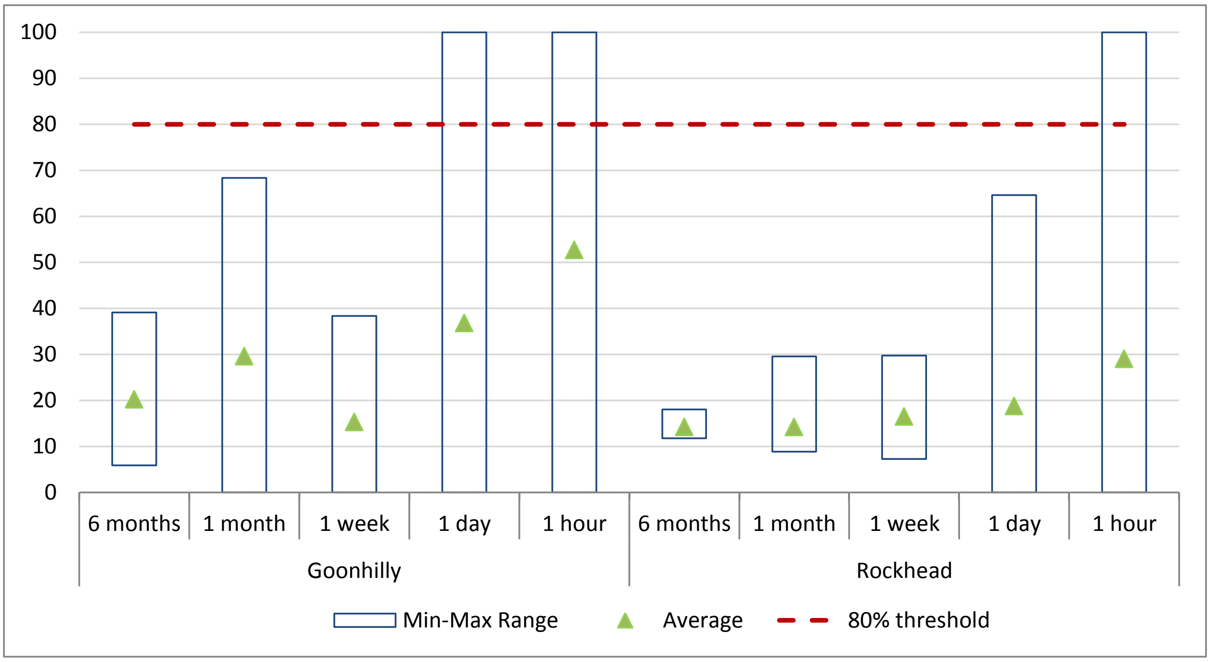
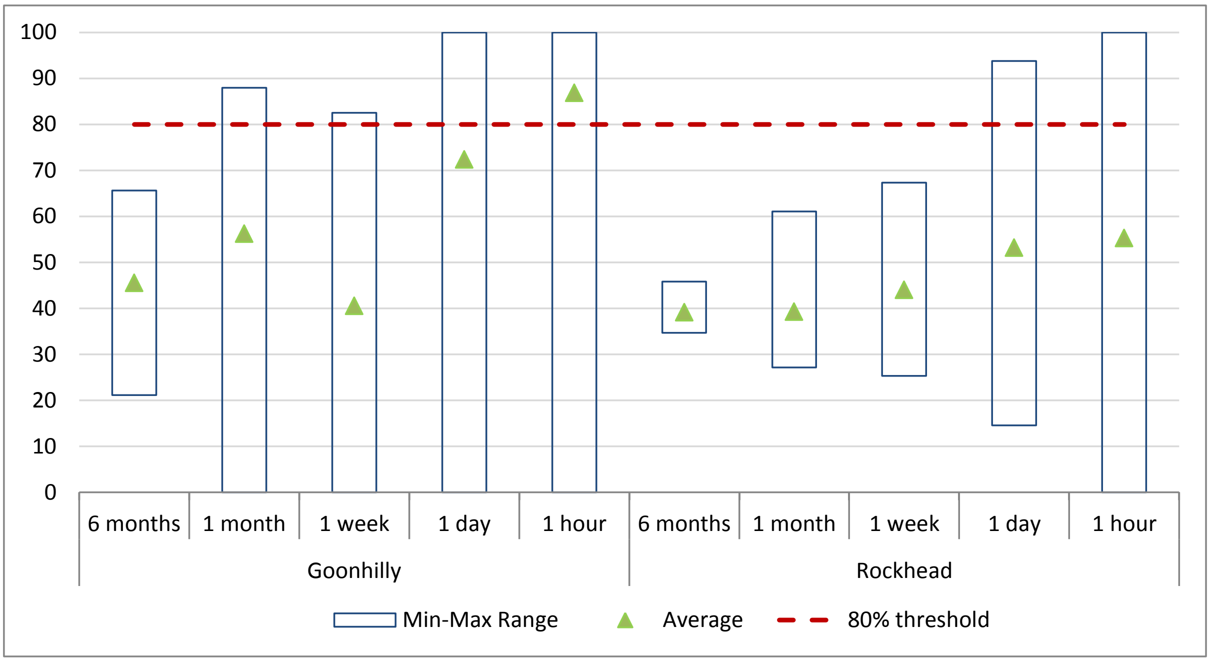


Figure 8. Percentage of predictions passing the 50% accuracy threshold



5.5. Load Customers

SCADA Data

Table 12. Testing Parameters for Load Customers – SCADA Data

Location	Time Horizons	Features	Tuning, Validation, Forecasting Periods	Data Inputs & Sources
Load Customer (SCADA) Jaguar Land Rover Wymeswold	Six Months Ahead Month Ahead Week Ahead Day Ahead Hour Ahead	Hour Day of Week Quarter Month Year Day of Year Day of Month Week of Year Holidays	Six Months Ahead: 17-12-2014 to 09-09-2017 6 simulations Month Ahead: 17-12-2014 to 01-11-2018 18 simulations Week Ahead: 17-12-2014 to 08-04-2018 18 simulations Day Ahead: 17-12-2014 to 15-12-2018 18 simulations Hour Ahead: 17-12-2014 to 09-02-2018 18 simulations	Jaguar Land Rover - T1 Incomer MW Jaguar Land Rover - T2 Incomer MW Wymeswold - Astrazeneca/Quorn MW Bank holidays for England and Wales

Table 13. Simulation Parameters for Load Customers – SCADA Data

Time horizon	Tuning period	Validation period	Training period	No. of simulations
Six Months Ahead	11 months	1 month	12 months	3-6
Month Ahead	11 months	1 month	12 months	7-18
Week Ahead	5 months	1 month	6 months	7-18
Day Ahead	5 months	1 month	6 months	7-18
Hour Ahead	5 months	1 month	6 months	7-18

NB: Lower number of simulations refers to Jaguar Land Rover, the higher number to Wymeswold

For large load customers, Capita DA has considered two data sources – SCADA data and Durabill data, as provided by WPD.

SCADA data contains continuous half-hourly readings in the same manner as for GSPs, BSPs, primaries and Generation Customers. Two customers were considered – Jaguar Land Rover and Wymeswold.

As an industrial consumer, Jaguar Land Rover follows a predictable load profile and can be modelled to exceed the 50% accuracy threshold for all time horizons. For Wymeswold, the general pattern is also identified by the model, however the peaks are not sufficiently accurate. Wymeswold accuracy is penalised by the MAPE error metric when readings are close to or equal to zero (which occurs for a significant proportion of the time for this particular customer).

Figure 9. Percentage of predictions passing the 80% accuracy threshold

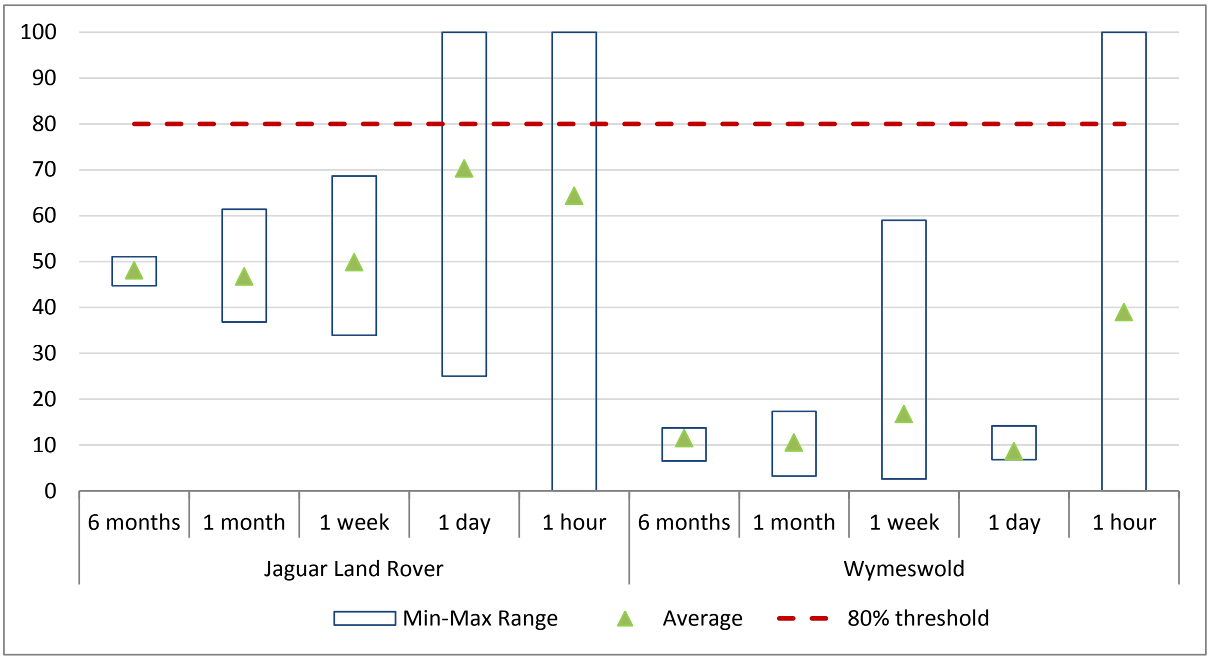
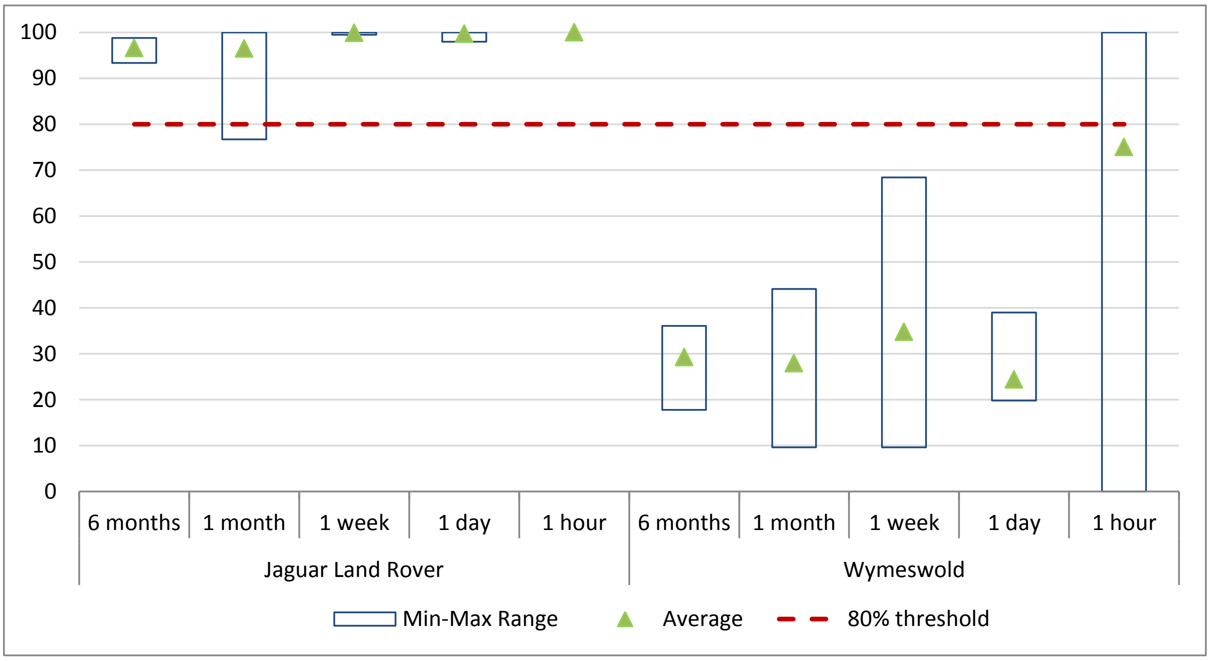


Figure 10. Percentage of predictions passing the 50% accuracy threshold



Durabill Data

Table 14. Testing Parameters for Load Customers – Durabill Data

Location	Time Horizons	Features	Tuning, Validation, Forecasting Periods	Data Inputs & Sources
Load Customer (Durabill) **14*00** **14*08** **14*05** **11*58** **11*59** **11*94**	Day Ahead Hour Ahead	Hour Day of Week Quarter Month Year Day of Year Day of Month Week of Year Holidays	14*00 03-05-2016 to 21-06-2016 3 simulations 14*08 08-10-2016 to 26-11-2016 3 simulations 14*05 29-10-2017 to 16-12-2017 3 simulations 1100039604358 14-04-2017 to 01-06-2017 3 simulations 11*59 24-10-2016 to 12-12-2016 3 simulations 11*94 23-11-2017 to 10-01-2018 3 simulations	MPANs: 14*00 14*08 14*05 11*58 11*59 11*94 Bank holidays for England and Wales

Table 15. Simulation Parameters for Load Customers – Durabill Data

Time horizon	Tuning period	Validation period	Training period	No. of simulations
Day Ahead	1 month	1 week	1 month and 1 week	3
Hour Ahead	1 month	1 week	1 month and 1 week	3

Durabill data also records half-hourly readings, however the readings in the data provided are not continuous and seem to be recorded only for specific time periods (e.g. one day in a week or on set dates for each month). This may reflect changes to the customer data within Durabill as customers categorised by Elexon as having profile classes 5-8 move to half-hourly settlement. Previous analysis using this data has not encountered issues with the completeness of datasets and it is likely that the customers that have traditionally been half-hourly metered (the 100kW market) would have good quality data. In terms of EFFS only the largest customers are likely to be connected at 33kV and above, which would exclude former profile class 5-8 customers. Capita DA has identified six customers with a sufficient body of data to allow for tuning, training and testing hour ahead and day ahead forecasts. In these cases the tuning and training periods were reduced, as per the simulation parameters above.

It is observed that all six entities passed the 80% accuracy threshold for hour ahead forecasts, and all six passed the 50% accuracy threshold for day ahead forecasts.

The key recommendation here would be to validate the assumption that those load customers that would require forecasts would have good quality, continuous data collection. As it stands, most cases considered lacked the continuous past data required for forecasting.

Figure 11. Percentage of predictions passing the 80% accuracy threshold

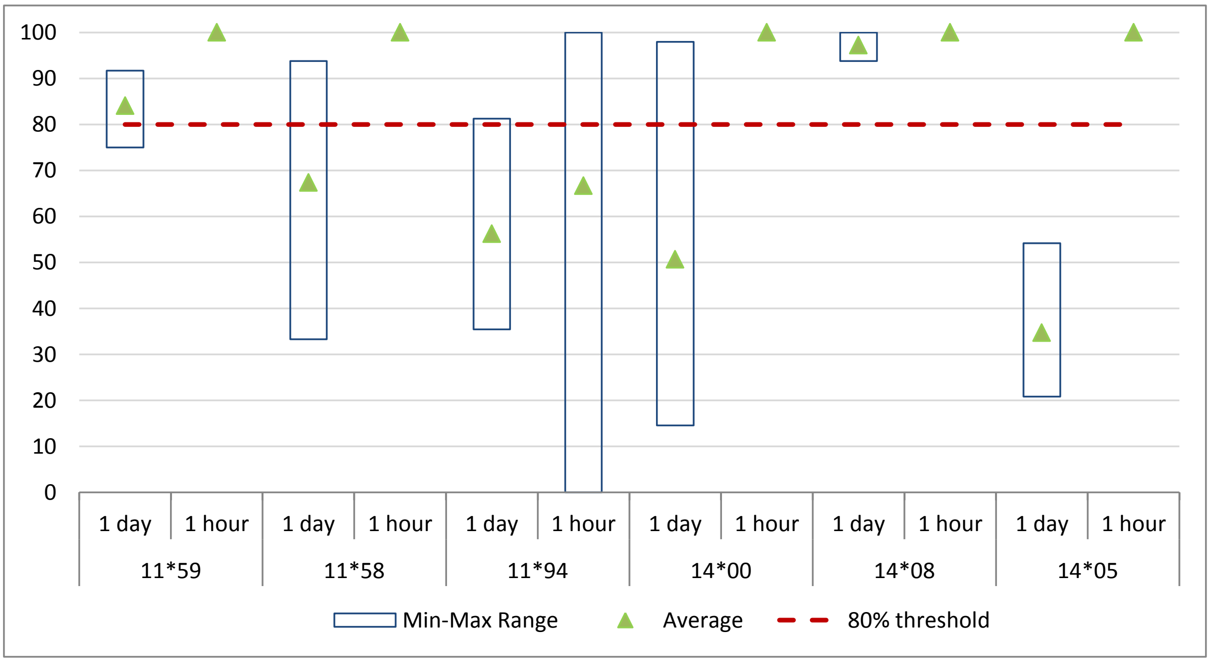
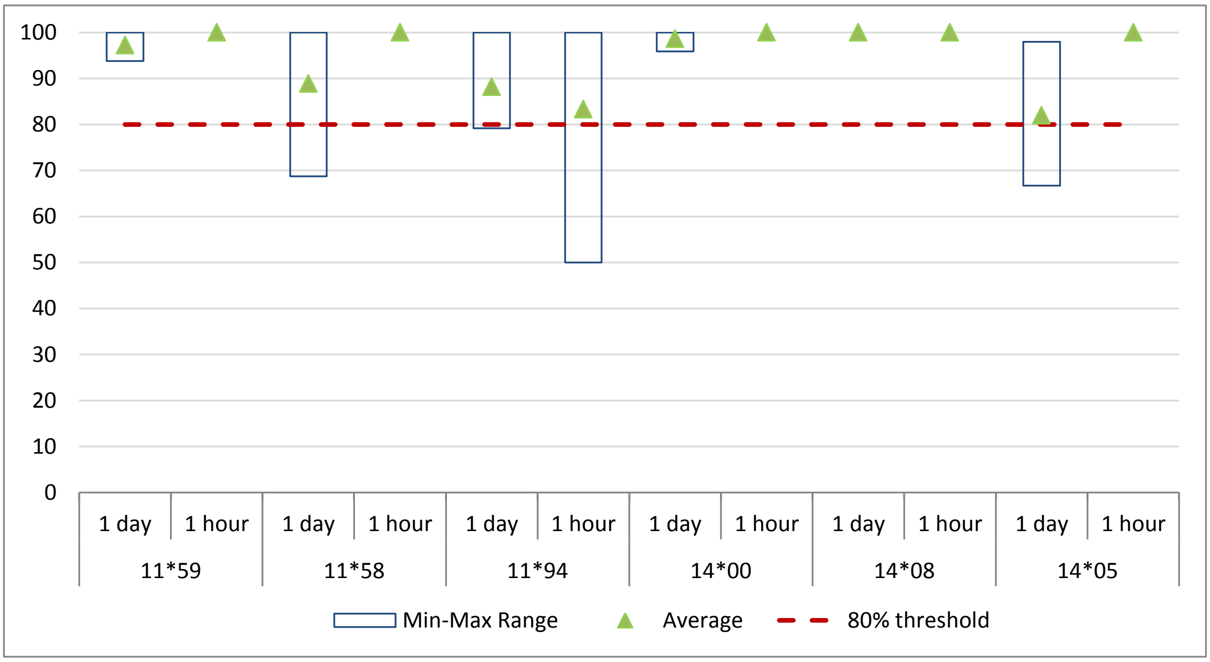


Figure 12. Percentage of predictions passing the 50% accuracy threshold



6. Comparison of Predicted vs. Actual Values

The following charts show a comparison of predicted and actual readings for a selection of simulations performed. For the locations used in testing, the charts show a best case, worst case and median case to help visualise the variation in results obtained and demonstrate the impact of factors such as data quality and predictability of the underlying behaviour.

The charts are designed to help the user visualise a number of aspects in the forecasting process, including:

- Evidence of issues with data quality;
- Observable patterns in the underlying behaviour;
- Quality of predictions
    - Where the model has performed well;
    - Where errors occur;
    - What range of prediction accuracy is observed for a set of simulations (i.e. the Best case, Worst Case and Median Case)

The most evident observation from these charts is that the poorest predictions are usually caused by data quality issues – for example where there is a clear offset between actual and predicted values, the cause is likely to be zero readings or error codes in the training data that bias the model. The recommendation here is to first check the training data for errors. If the errors occur in one part of the dataset, it may be possible to shorten the training history in order to train the model on the valid data only. Examples of evidence of data quality issues are:

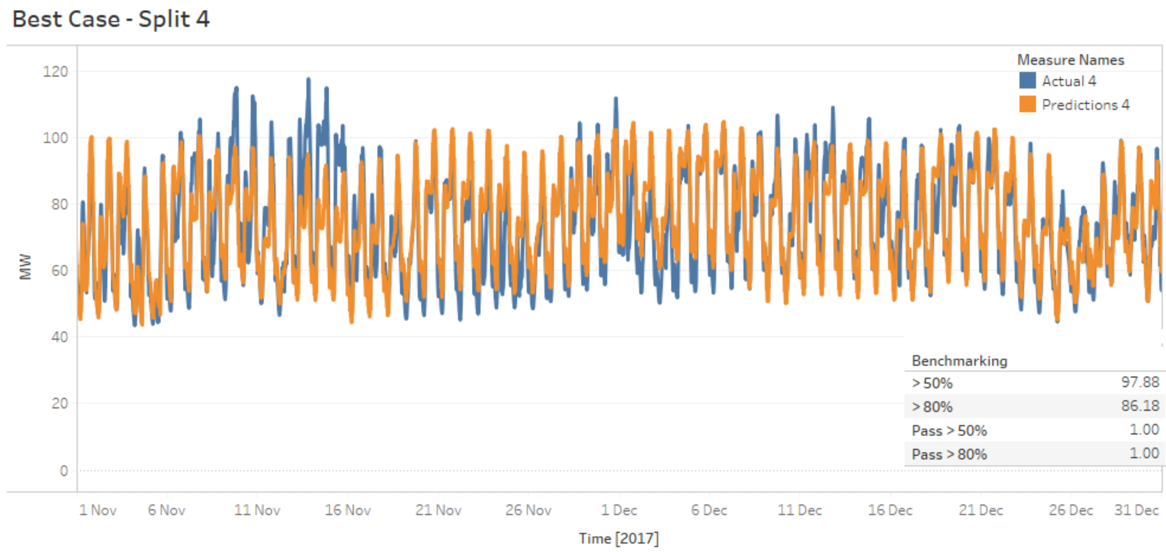
- Periods of zero readings in GSP Landulph TX2 and TX4 bias the model;
- Spikes of negative readings (possibly error codes) in BSP Morriston;
- BSP St Clears shows incorrect readings throughout the data;

Once data quality is excluded, some general observations can be made:

- For GSPs, the models are able to extract a general pattern and daily range, with errors occurring in the magnitude of daily peaks and changes in trend;
- For BSPs and primaries where the acceptance criteria are reached, a clear and stable pattern can be observed that is correctly predicted by the model;
- For BSPs and primaries where the acceptance criteria are not reached, model performance suffers where there is a change of pattern or direction (e.g. the daily range of load shifts up or down). Exploring with feature optimisation may help improve performance, e.g. if the directional change is driven by renewable generation – the same can be explored with GSPs, keeping in mind that they cover a larger area;
- For wind farms, longer-term patterns are difficult to observe and the models are only able to predict short-term time horizons. Optimising the feature set to remove some of the temporal features may improve accuracy somewhat;
- Load customers vary between them, cases where the behaviour is stable and cyclical will be reasonably well predicted by the model, with errors occurring again in the magnitude of daily peaks and changes in direction.

6.1. GSPs

Figure 13. GSP – Landulph – TX4 – six months ahead



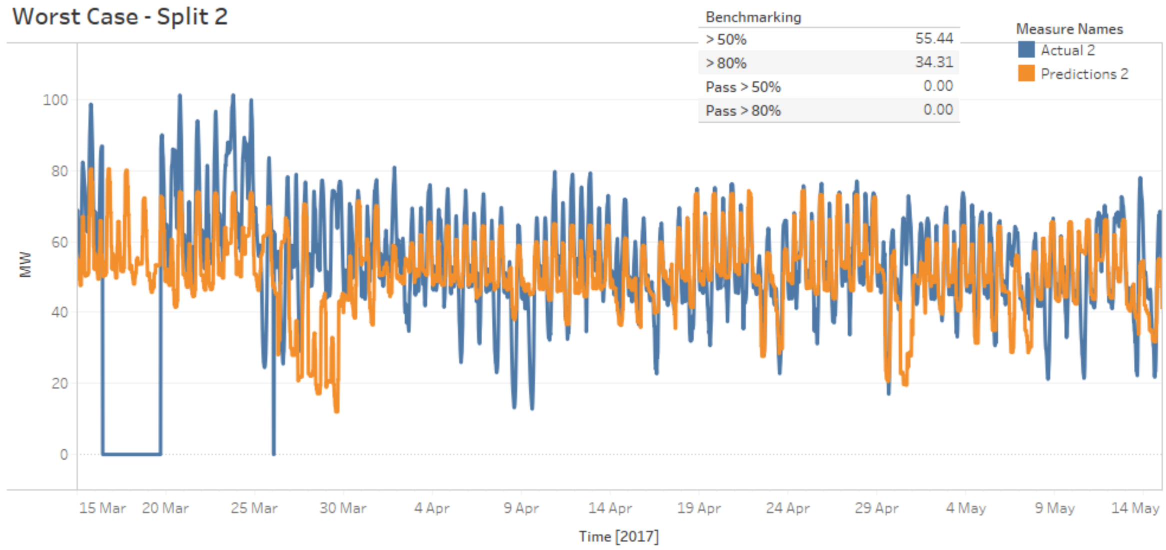
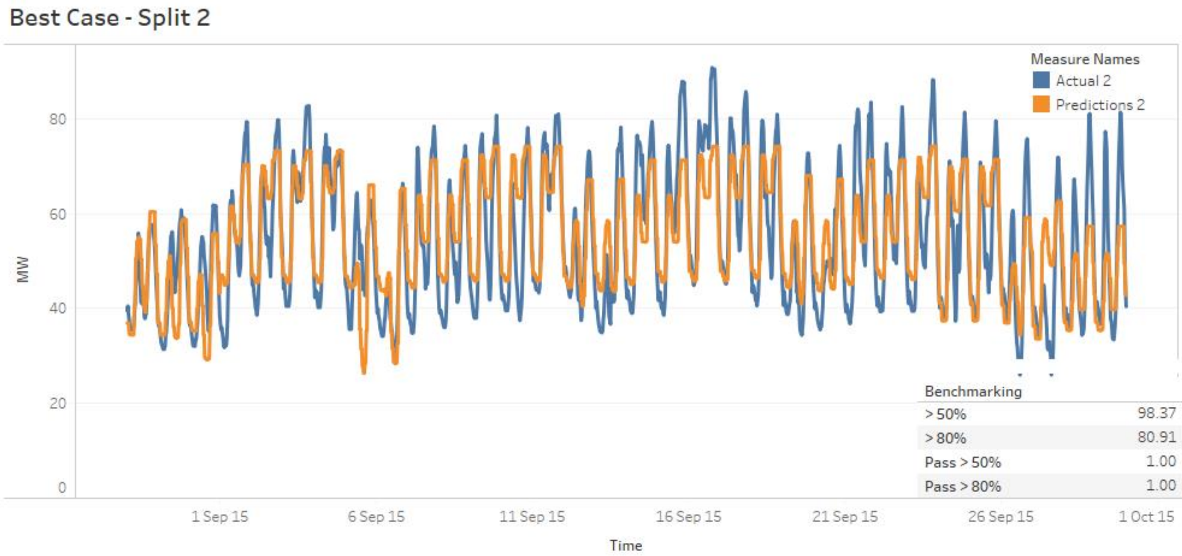
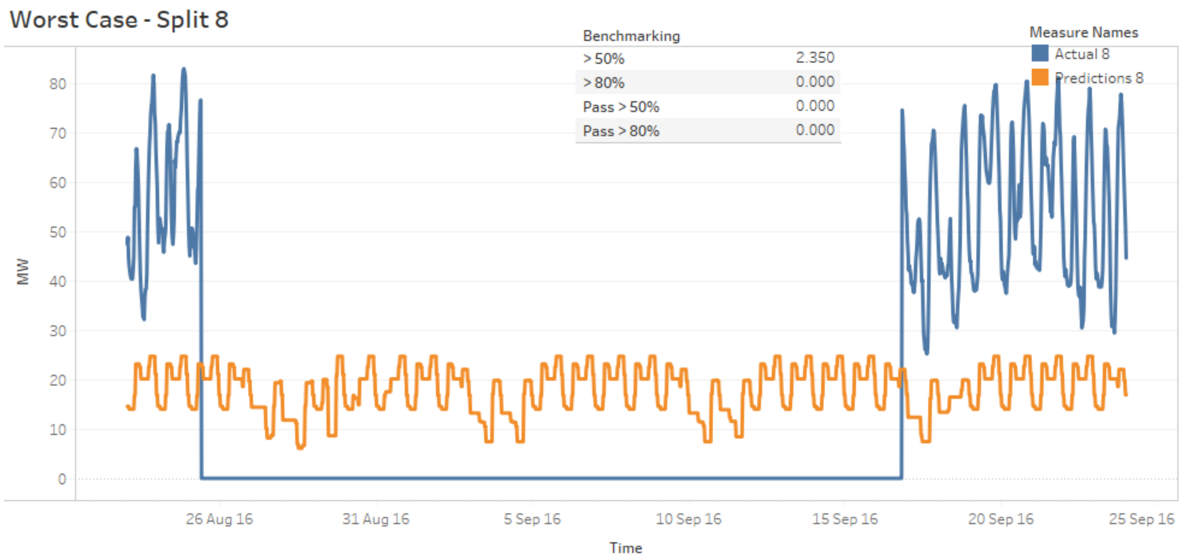


Figure 14. GSP – Landulph TX2 – month ahead





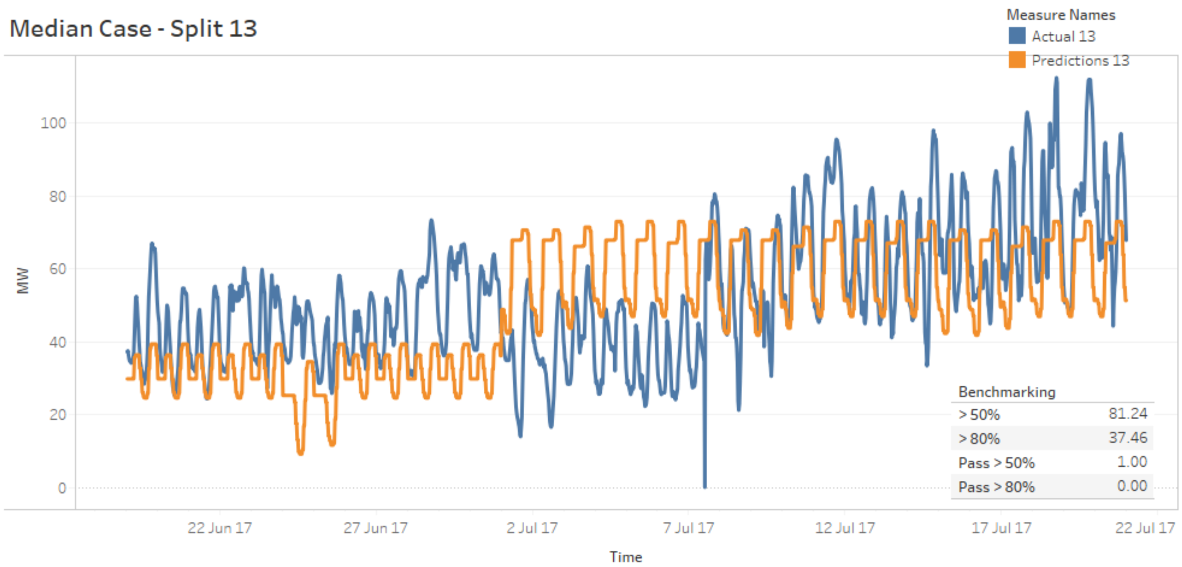
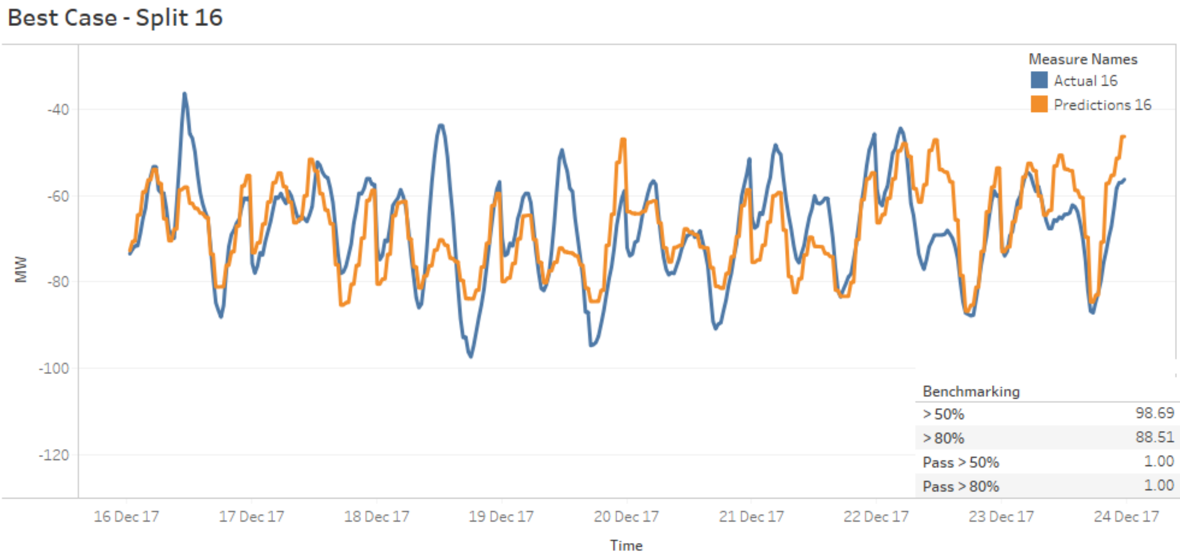
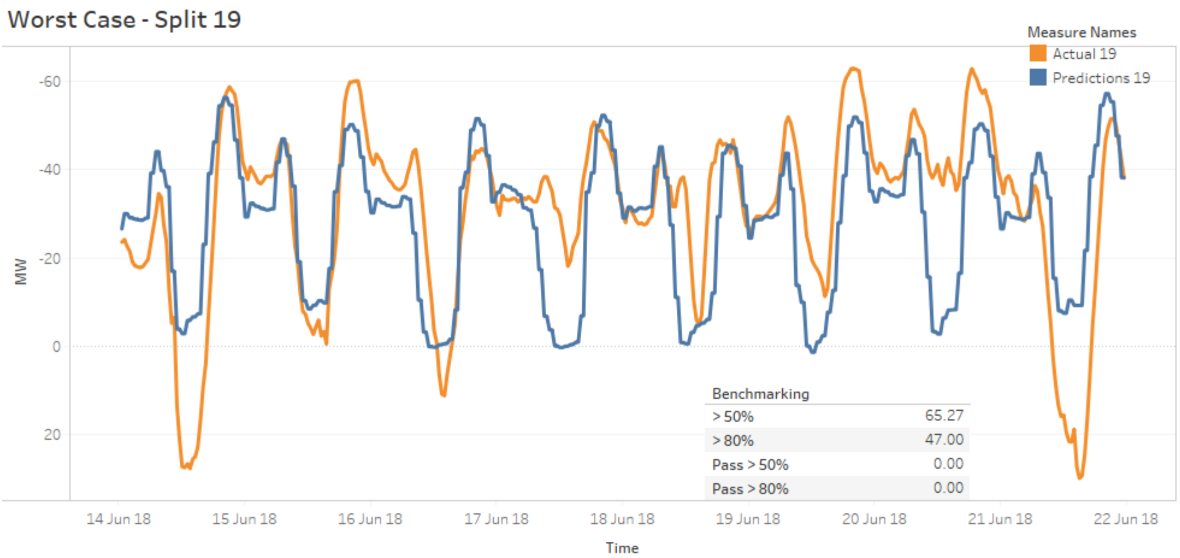


Figure 15. GSP – Indian Queens – TX4 – week ahead





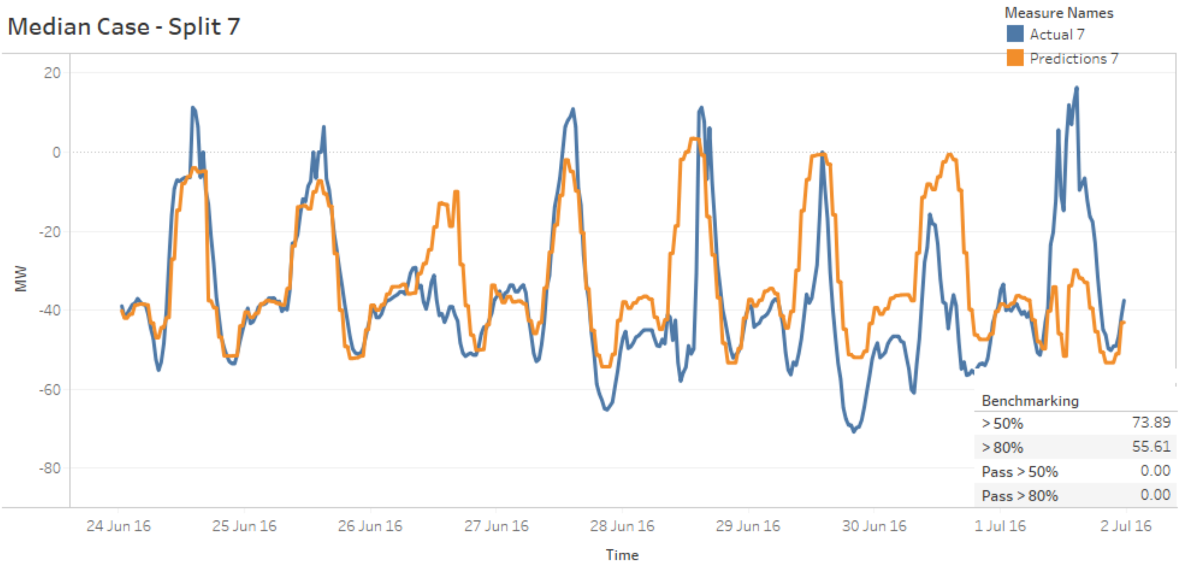
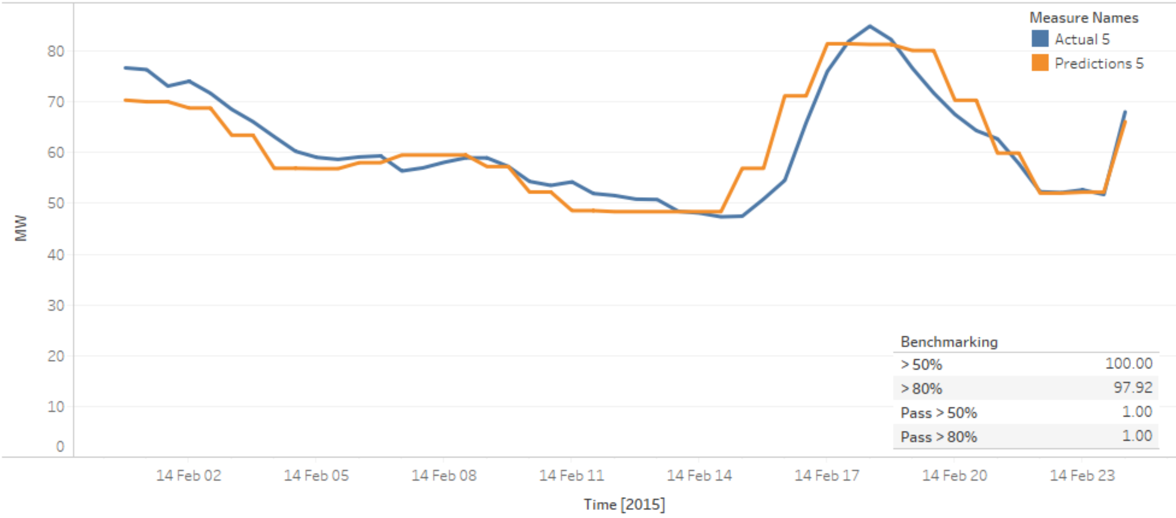
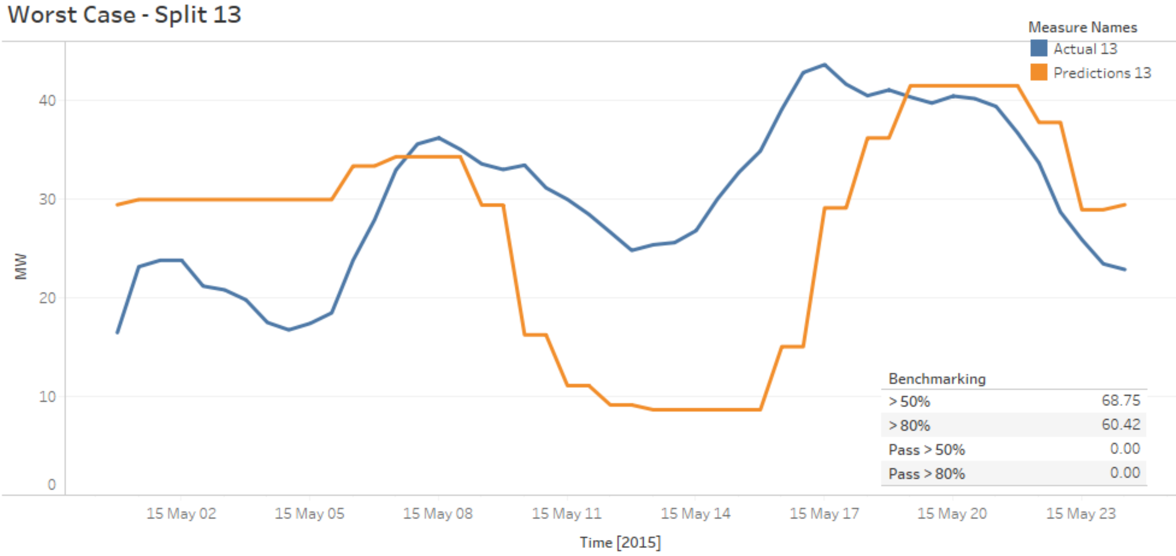


Figure 16. GSP – Indian Queens – TX3 – 1 day

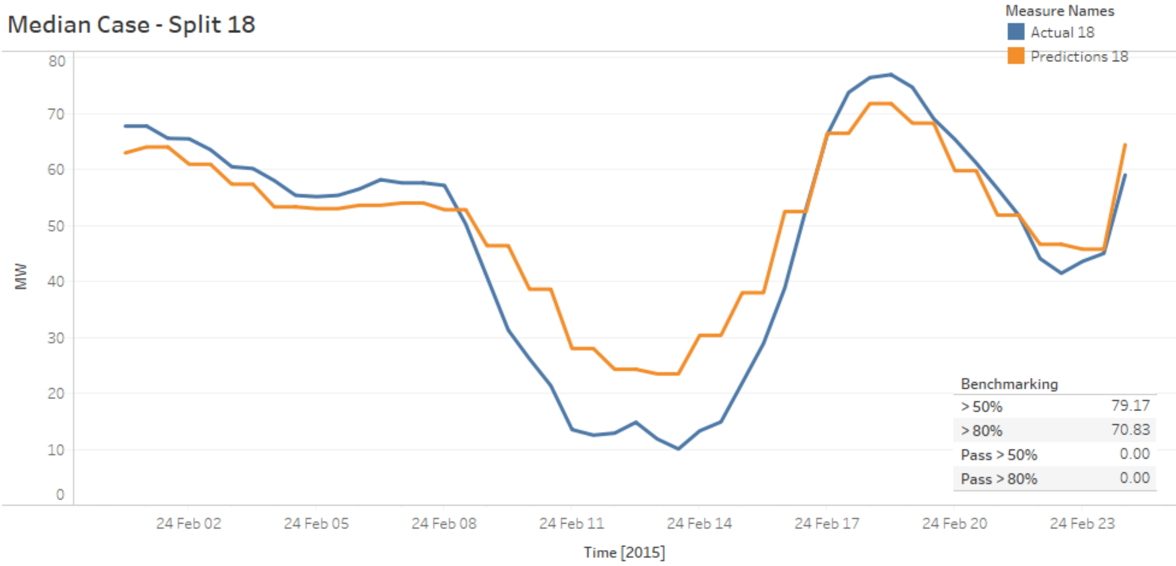
Best Case - Split 5



Worst Case - Split 13

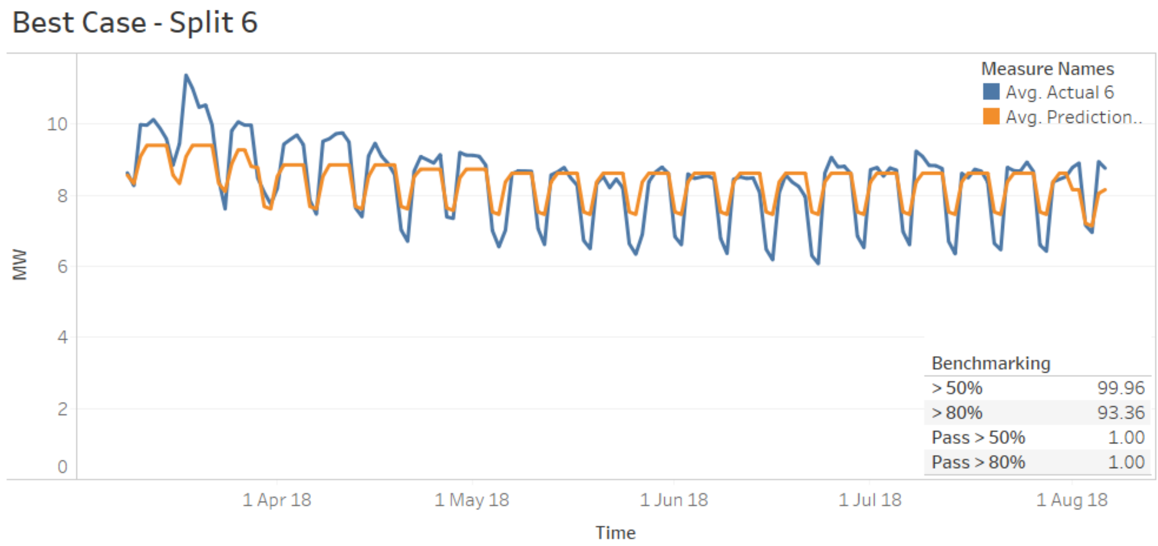


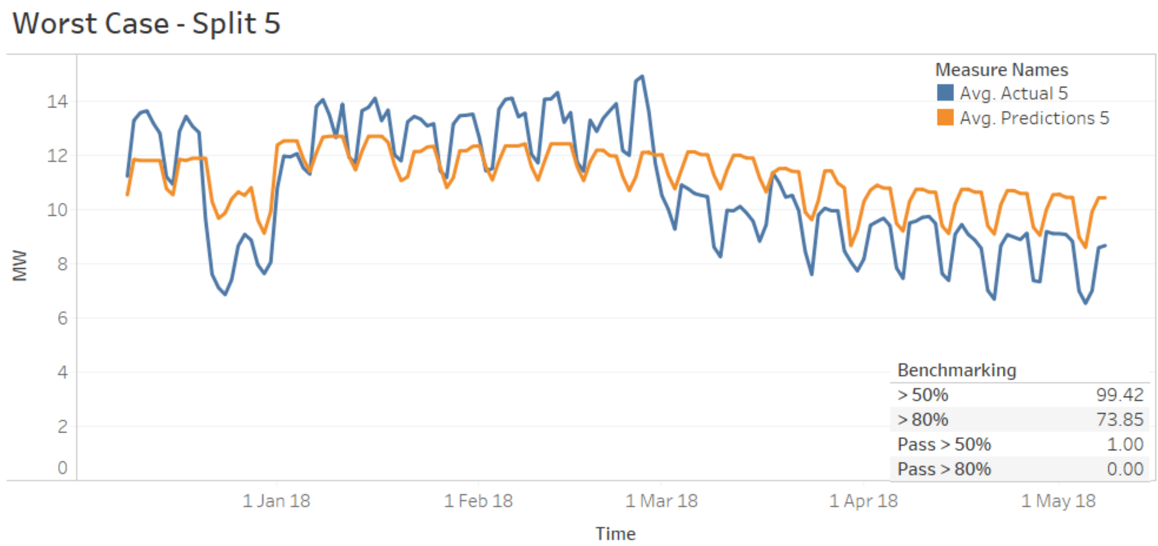
Median Case - Split 18



6.2. BSPs

Figure 17. BSP - Cardiff South - 6 months ahead





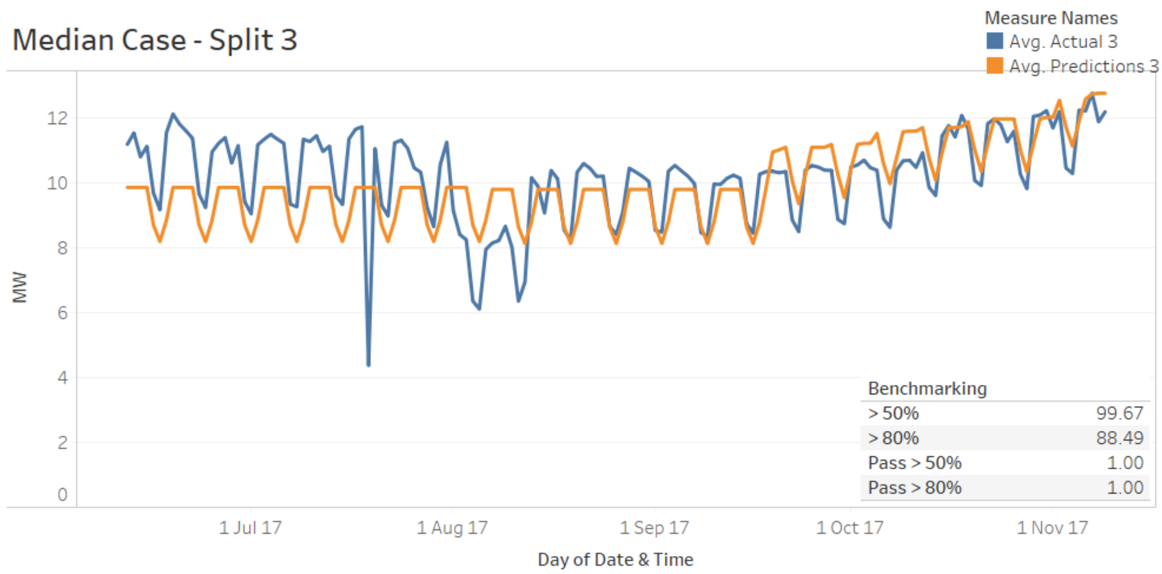
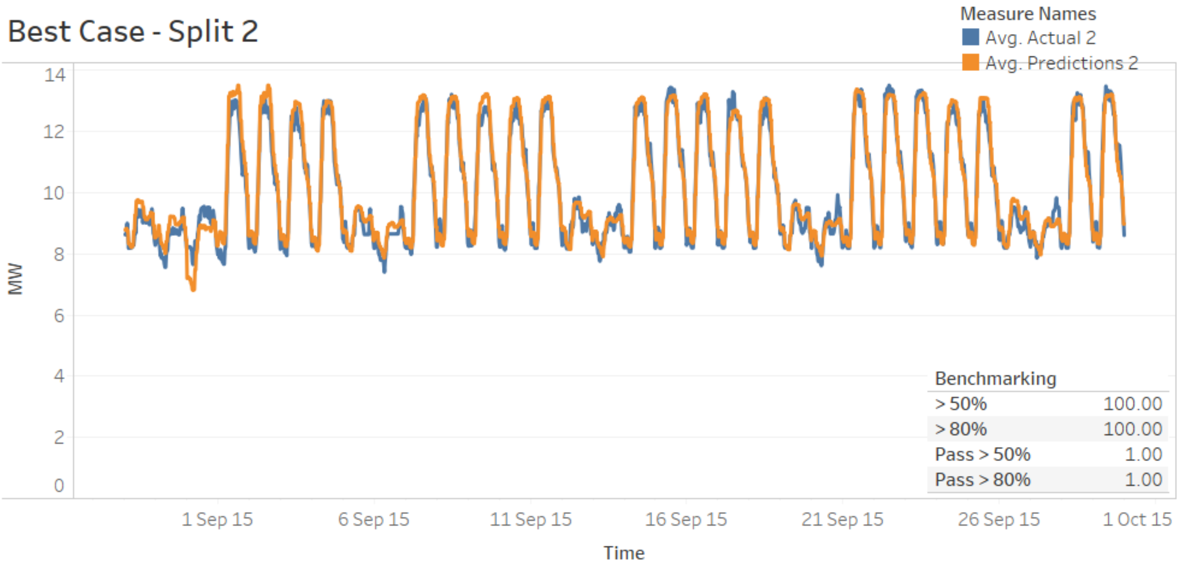
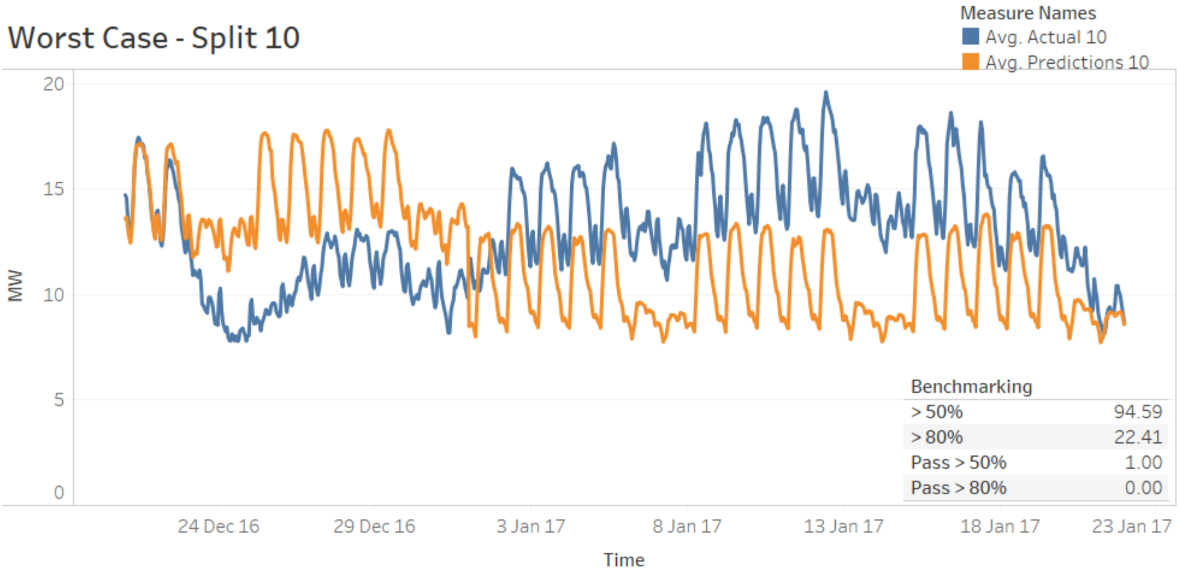


Figure 18. BSP - Cardiff South – 1 month ahead





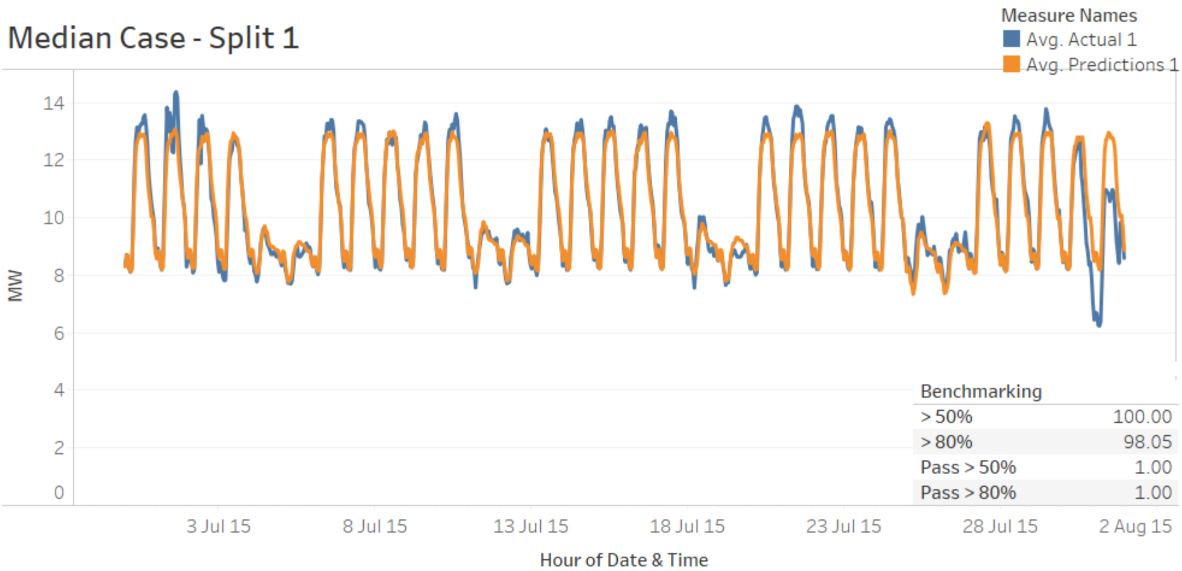


Figure 19. BSP - Morriston – 1 month ahead

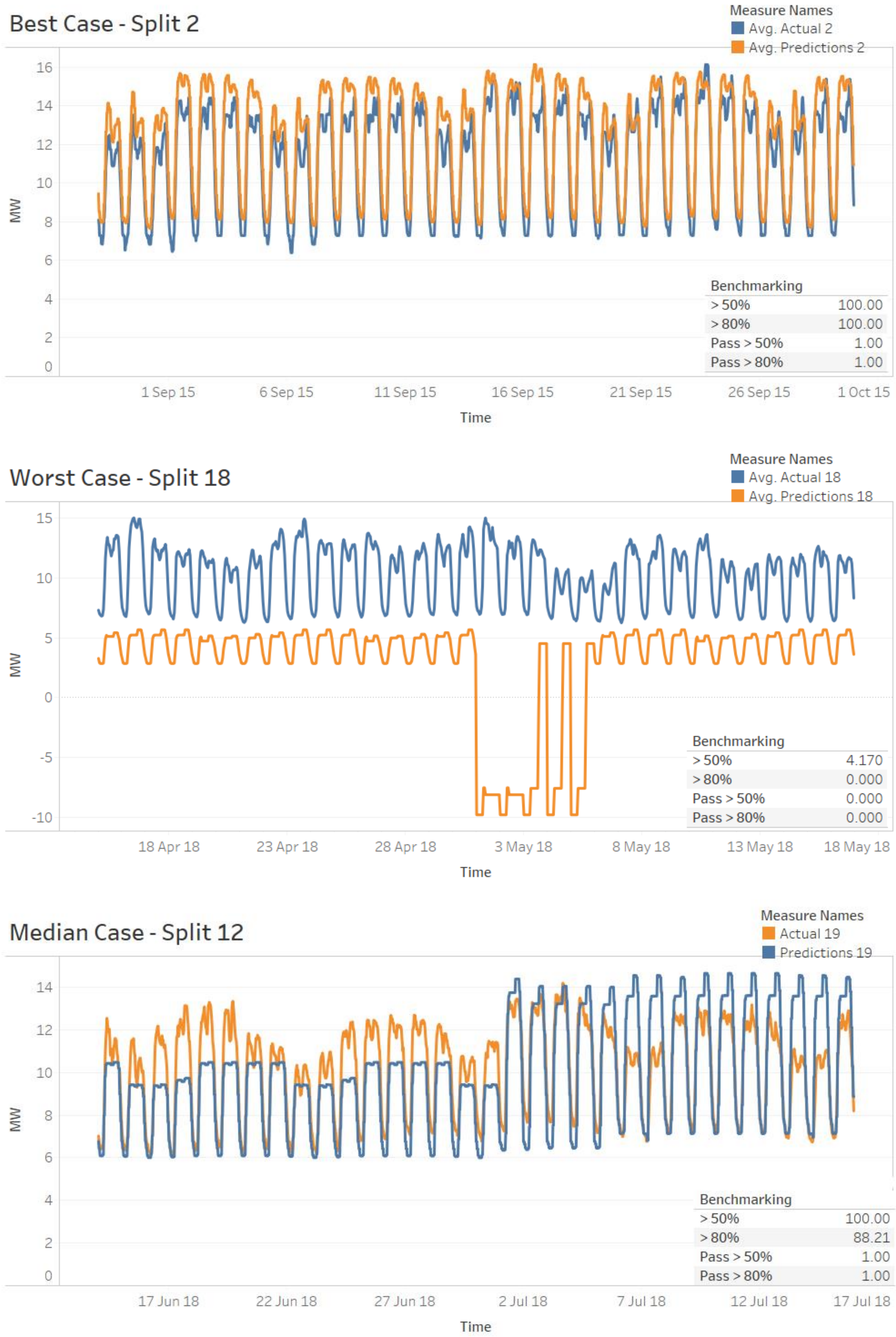
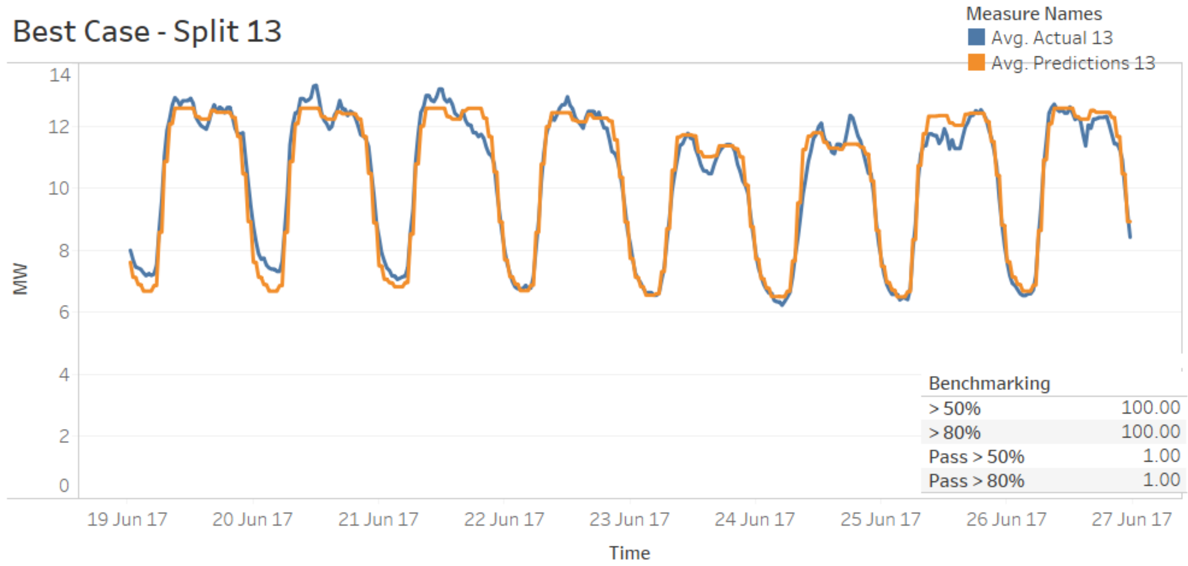
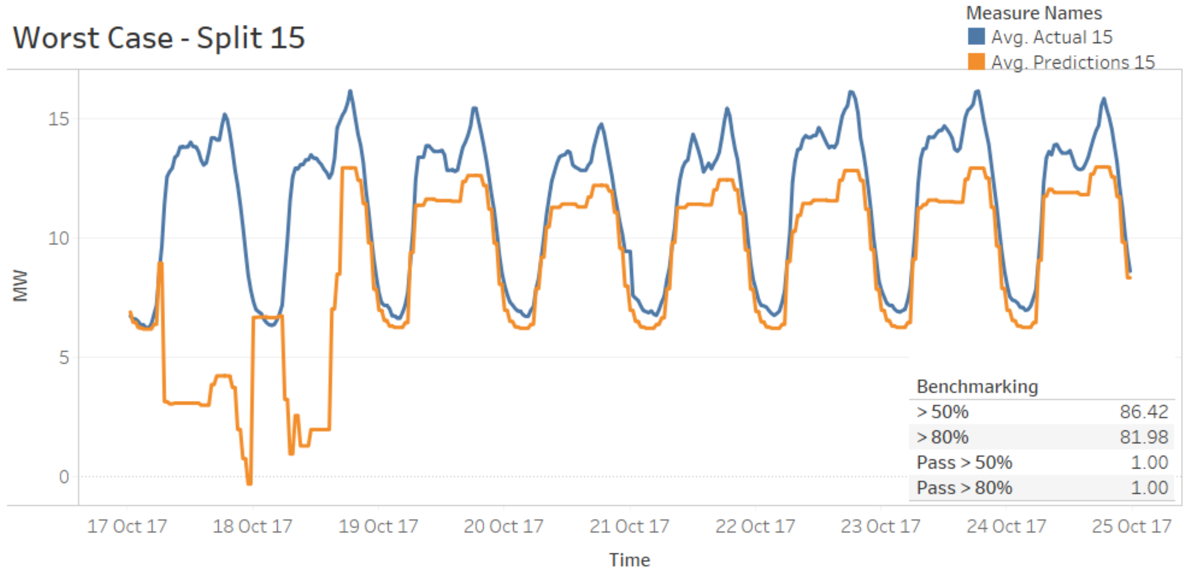
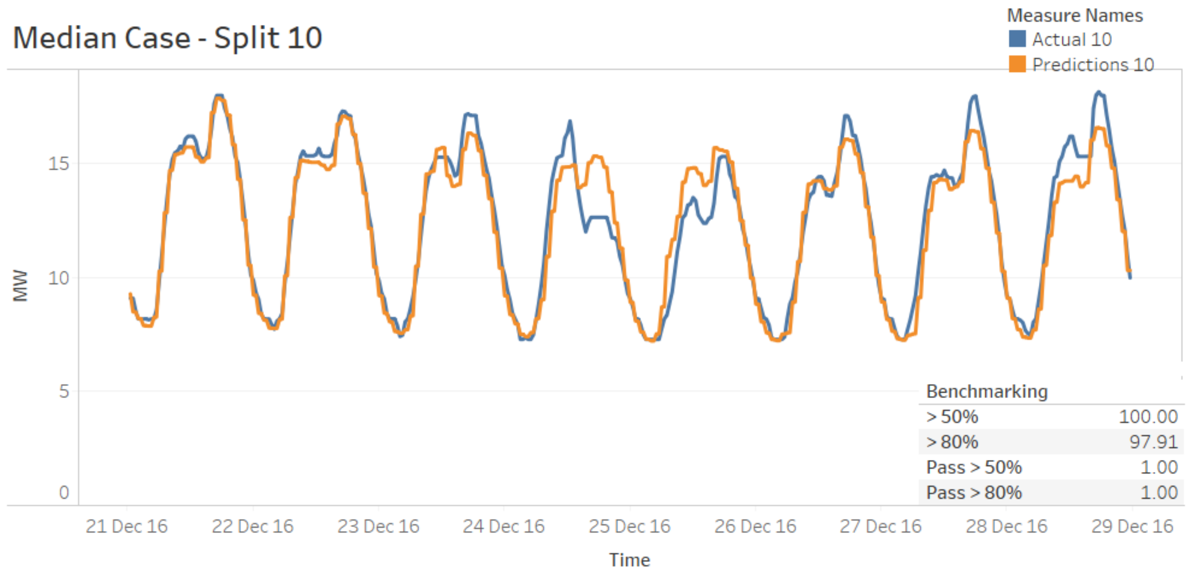


Figure 20. BSP – Morriston – 1 week ahead

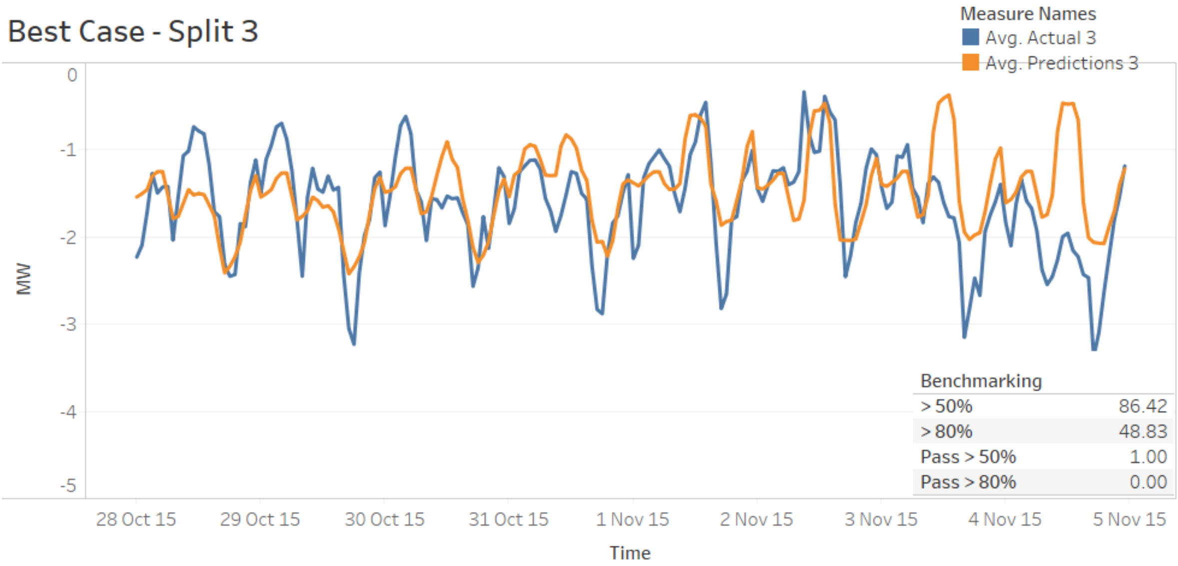






WESTERN POWER DISTRIBUTION
EFFS

Figure 21. BSP - Truro – 1 week ahead



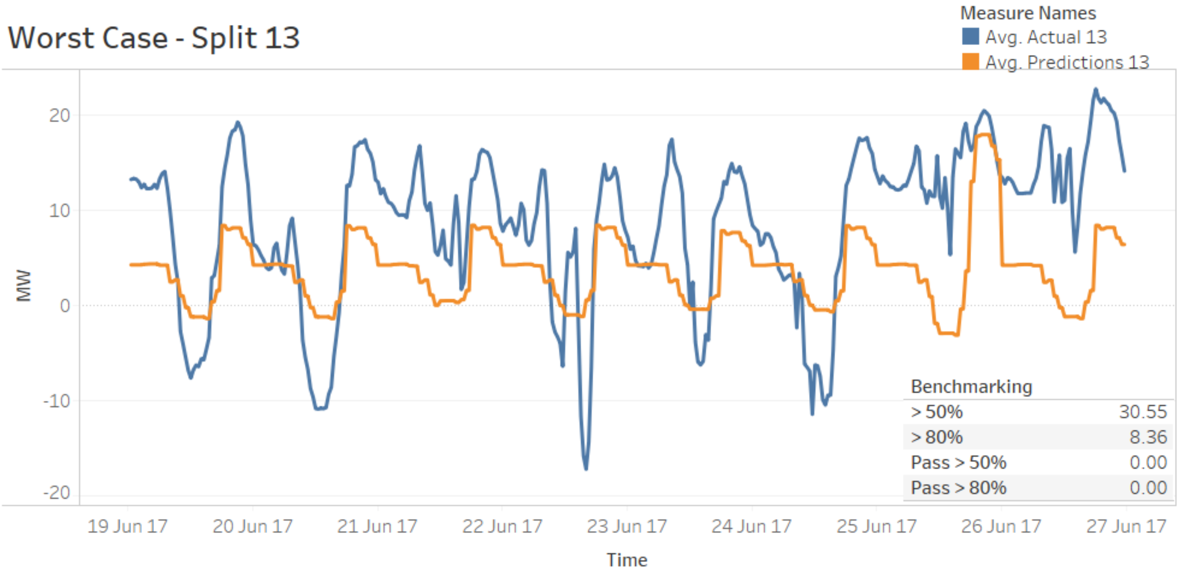
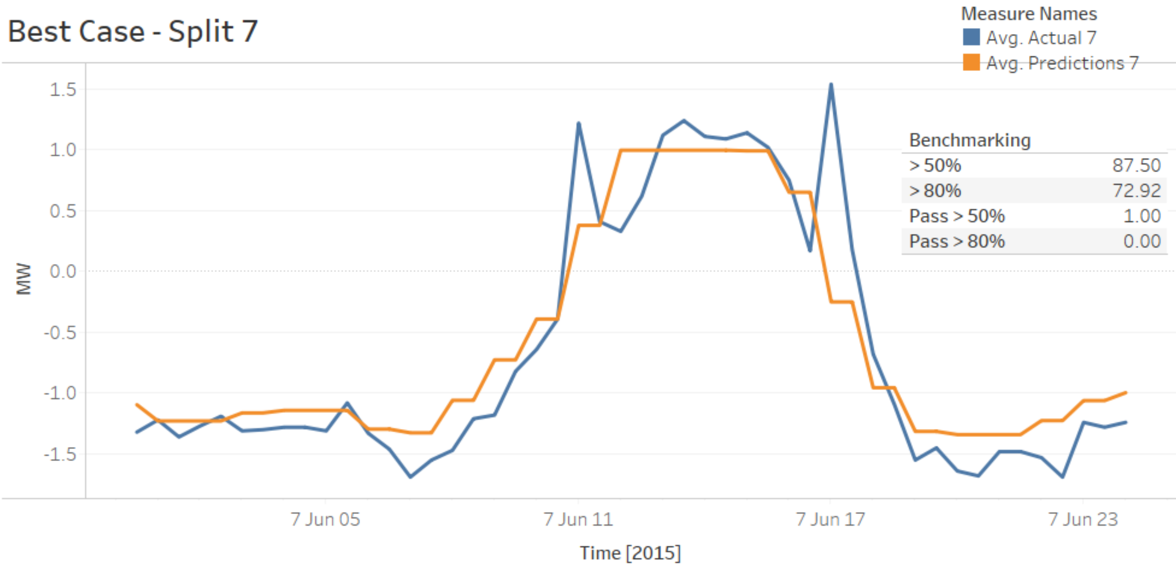
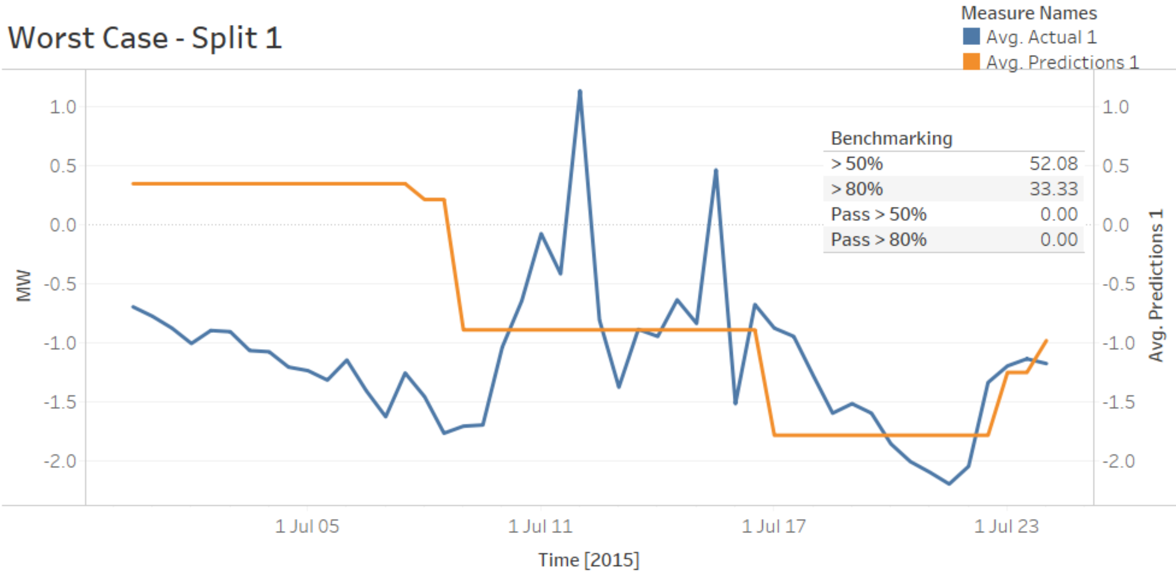


Figure 22. BSP - Truro – 1 day ahead





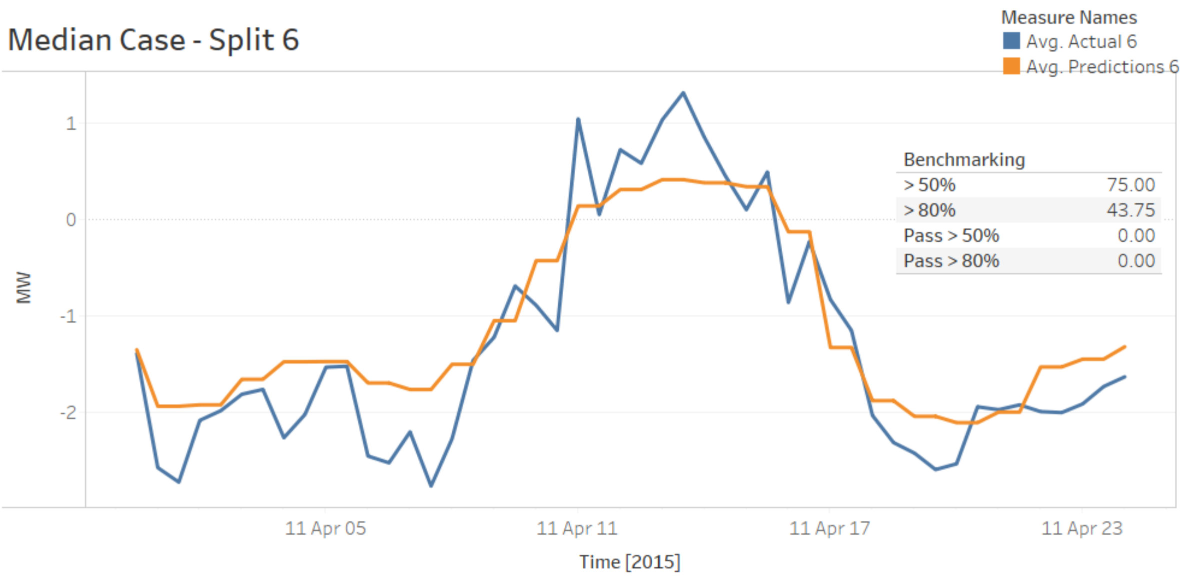
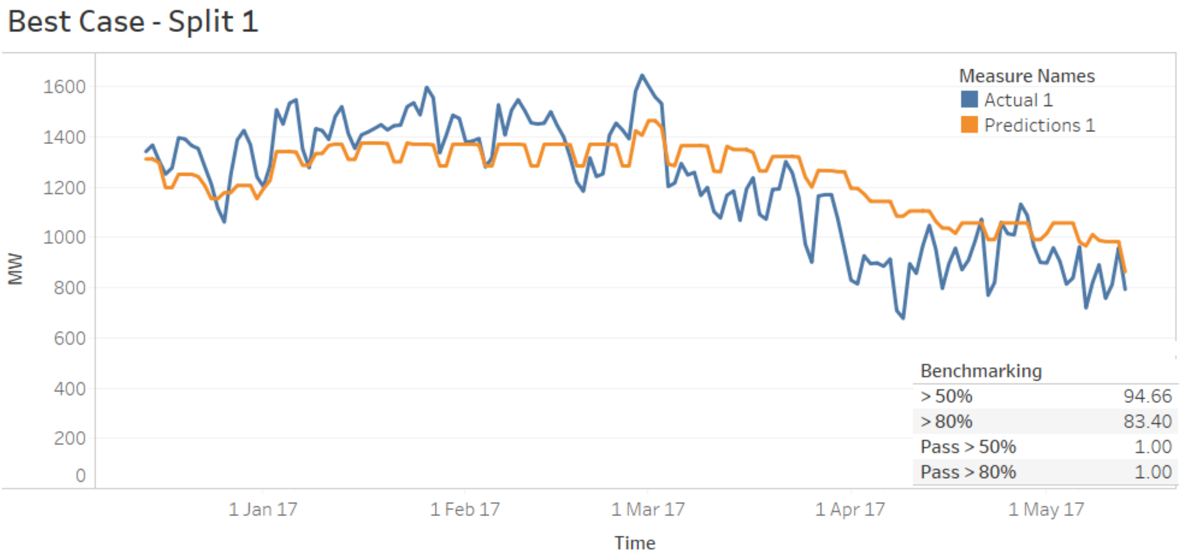
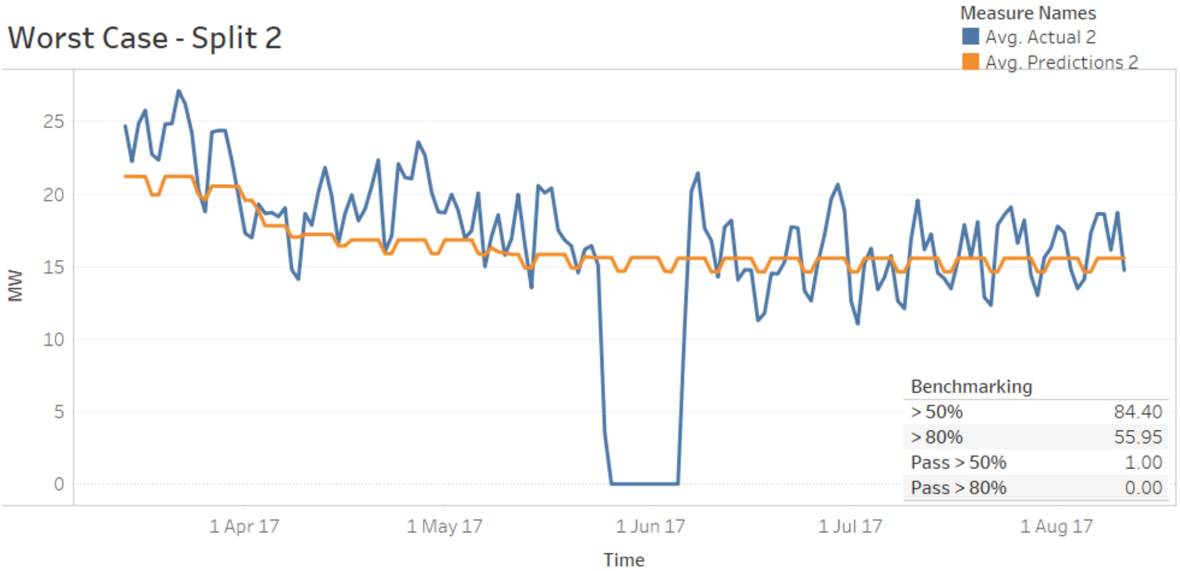


Figure 23. BSP - Ludlow – 6 months





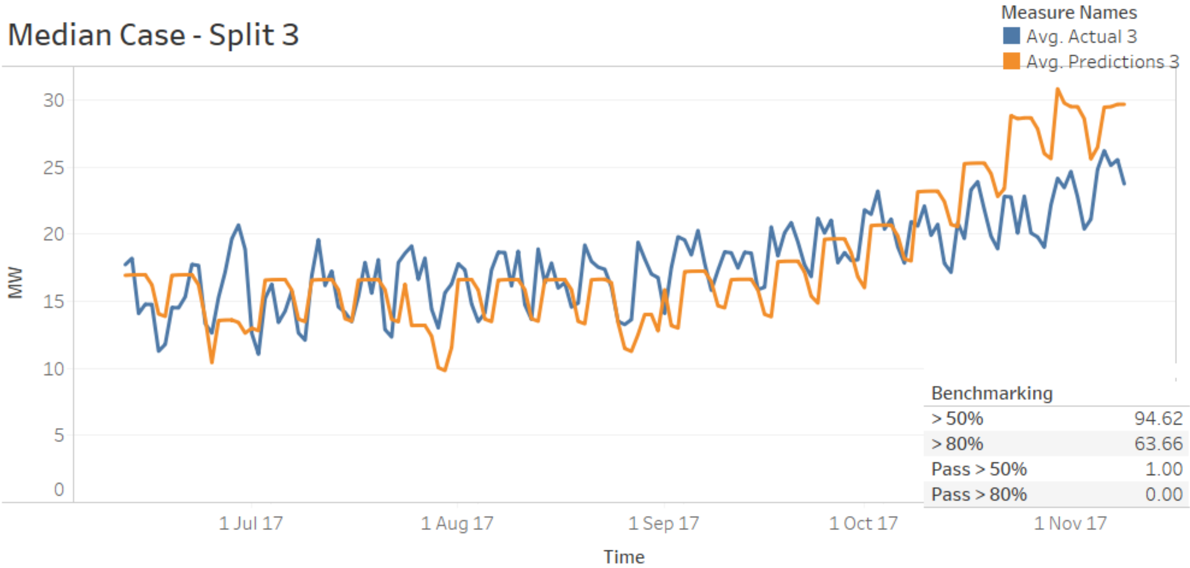
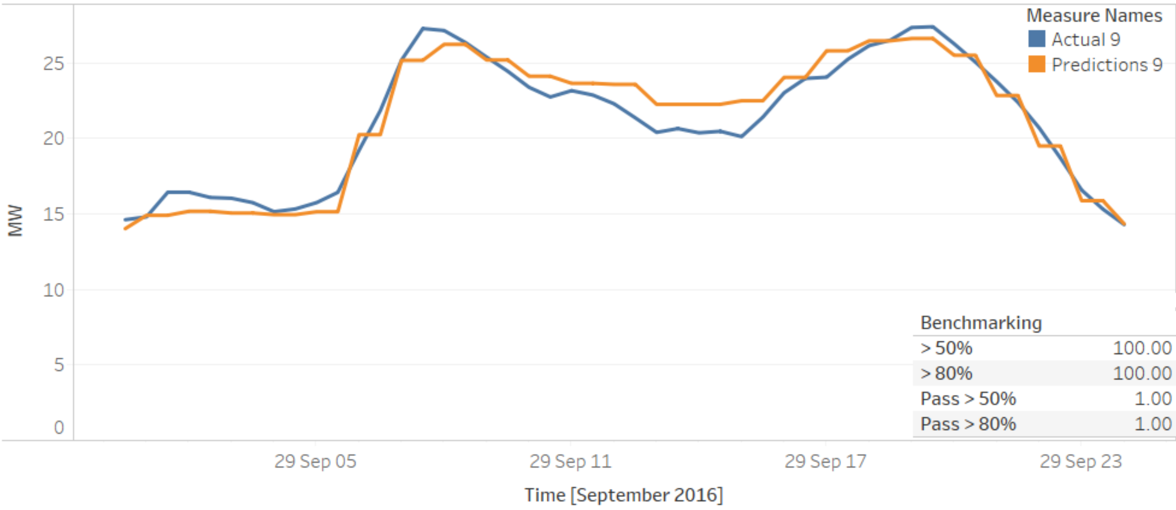
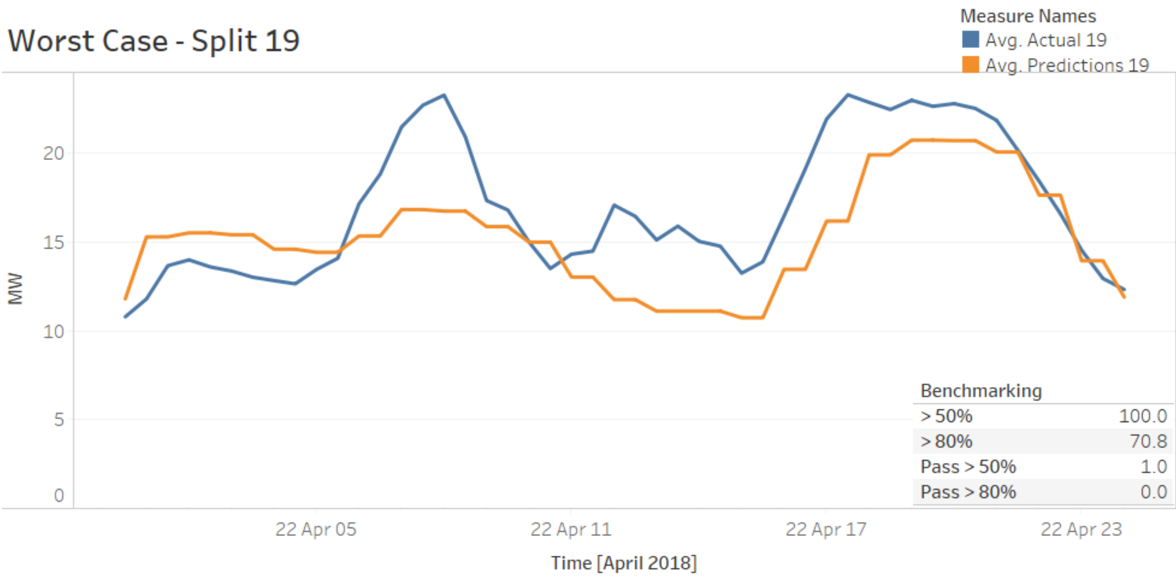


Figure 24. BSP - Ludlow – 1 day

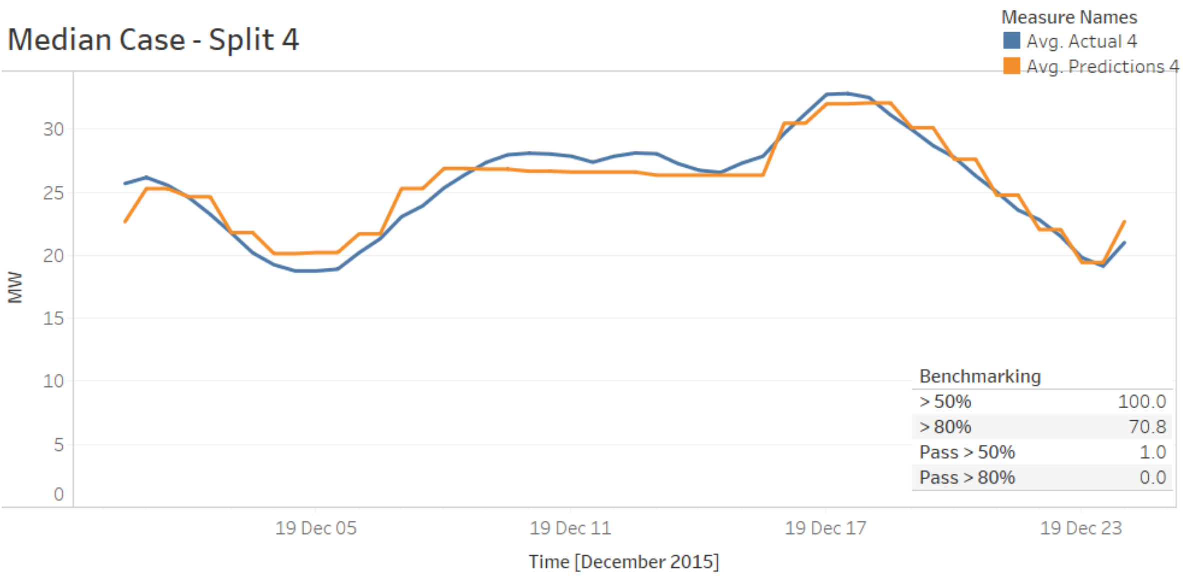
Best Case - Split 9



Worst Case - Split 19

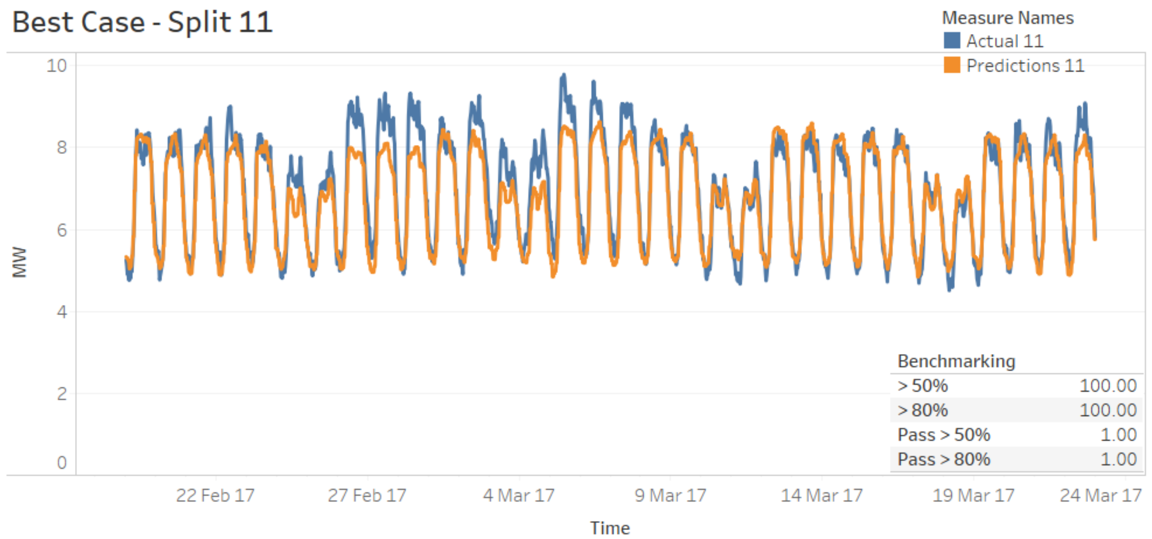


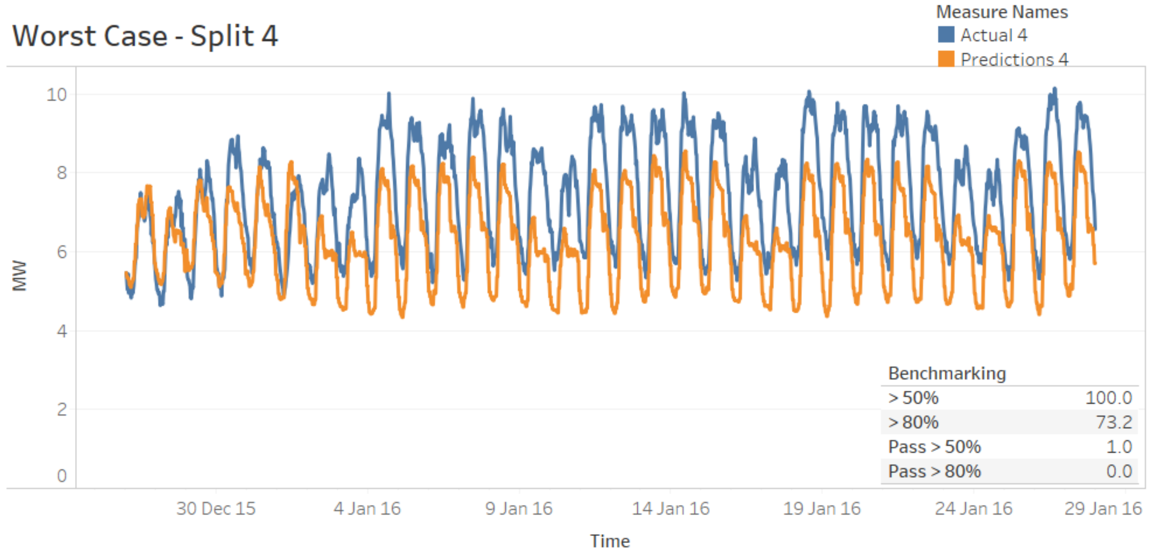
Median Case - Split 4



6.3. Primaries

Figure 25. Primary – Prince Rock – 1 month





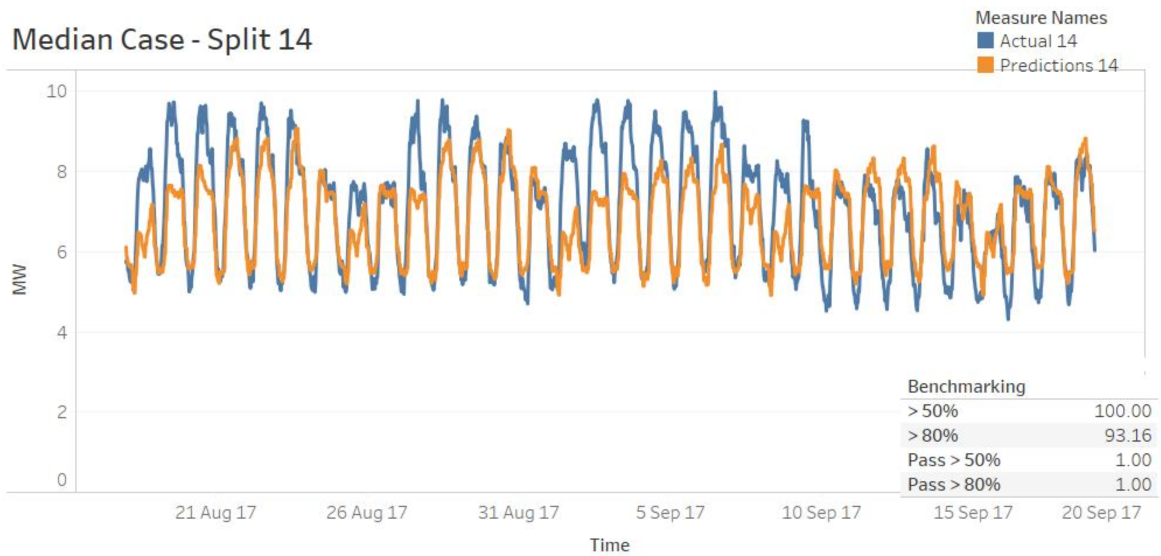
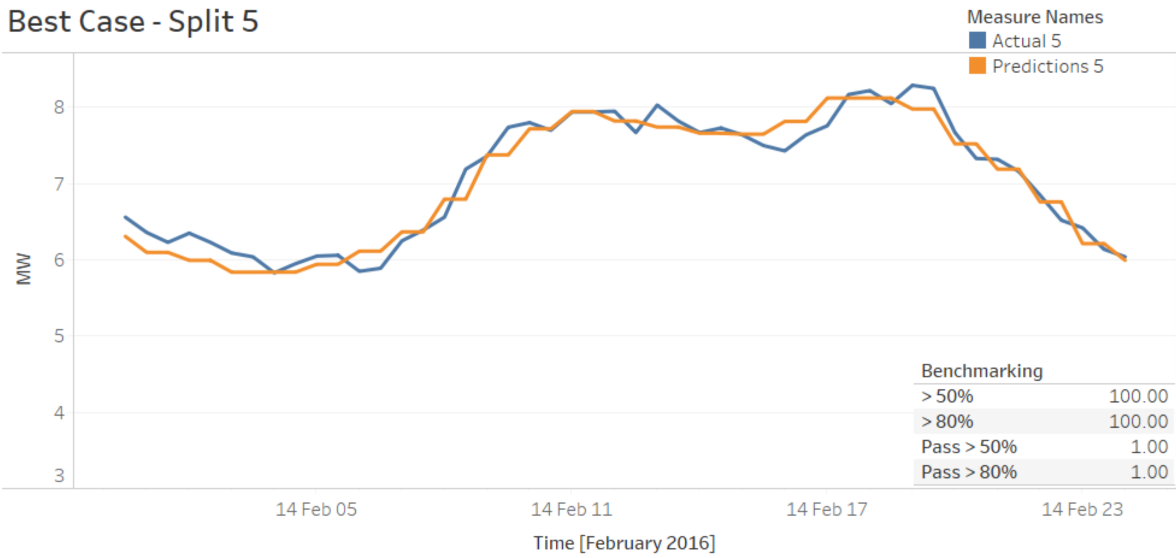
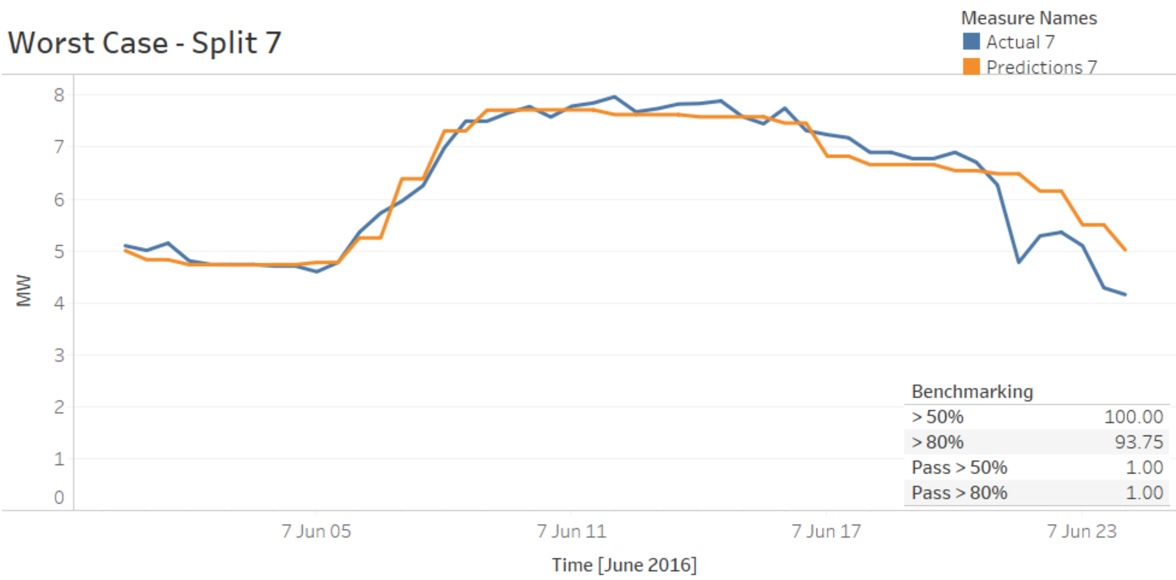


Figure 26. Primary – Prince Rock – 1 day





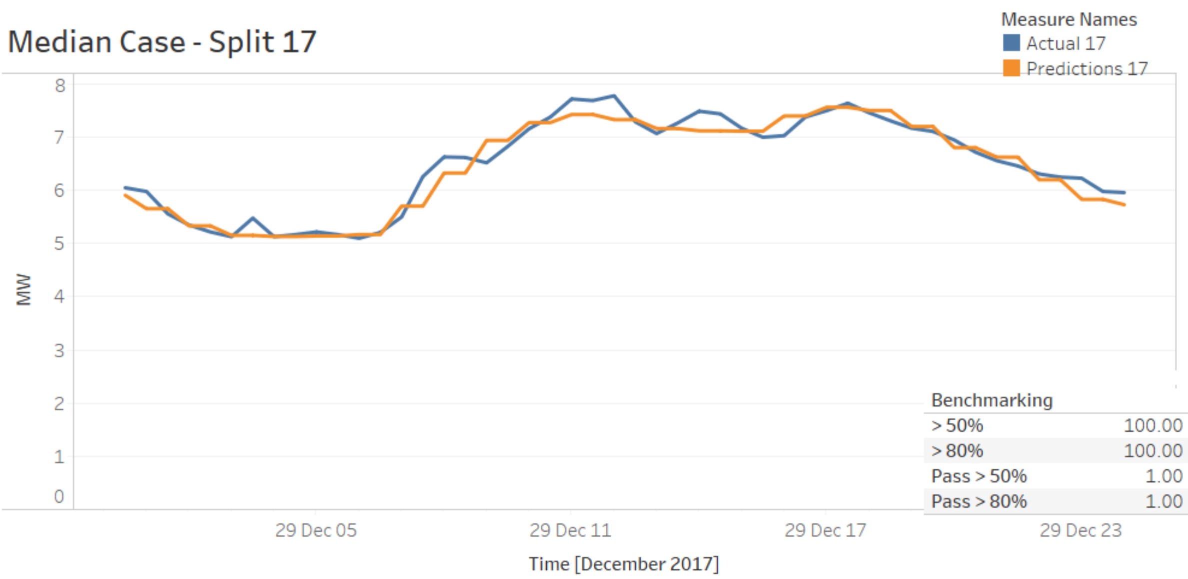
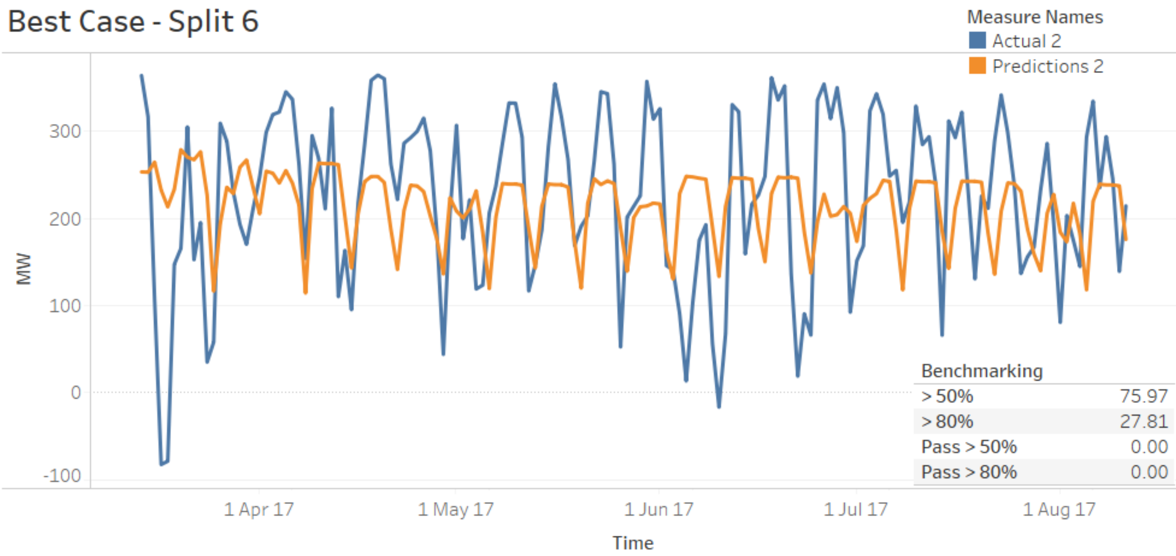
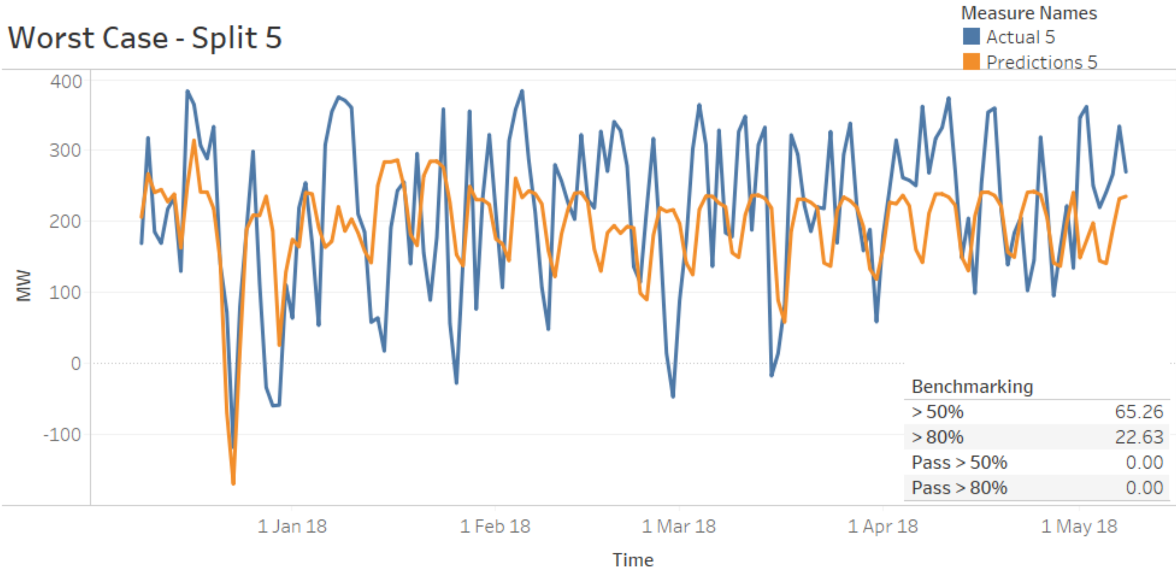


Figure 27. Primary – Kingsweston – 6 months





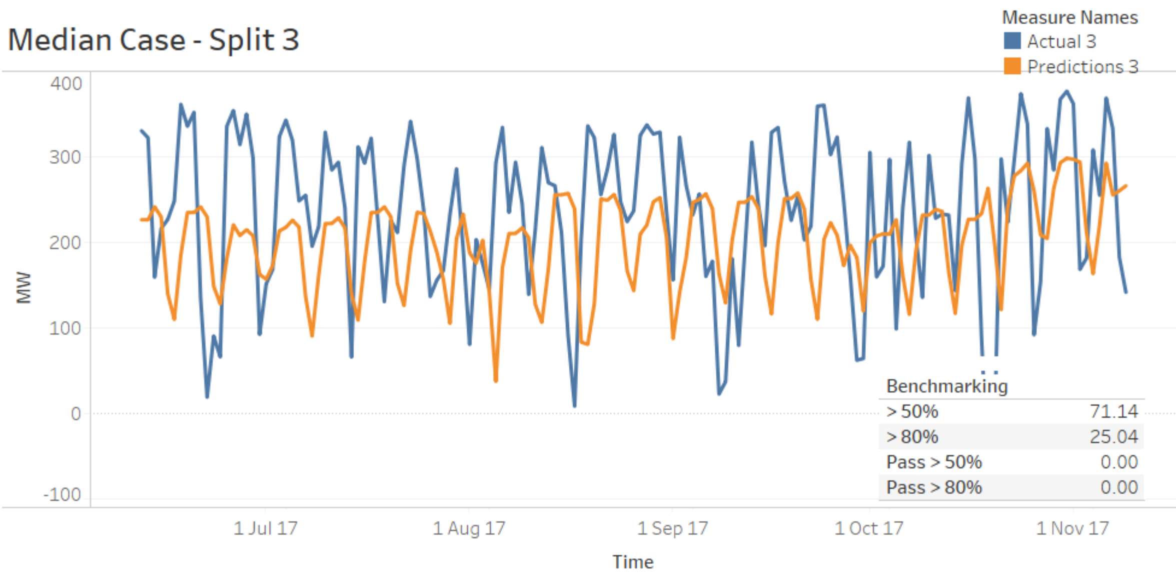
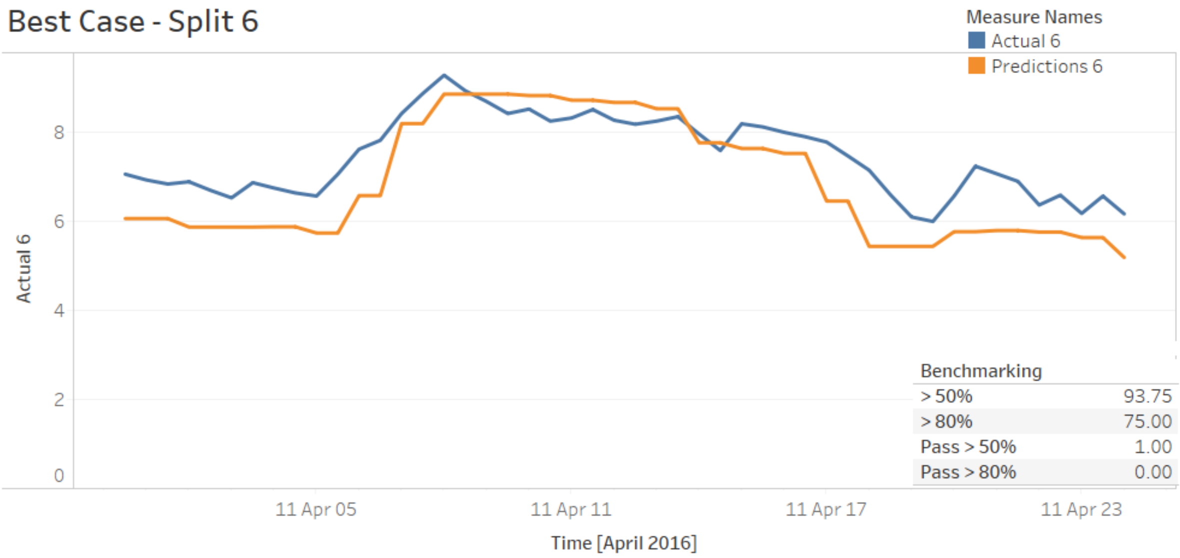
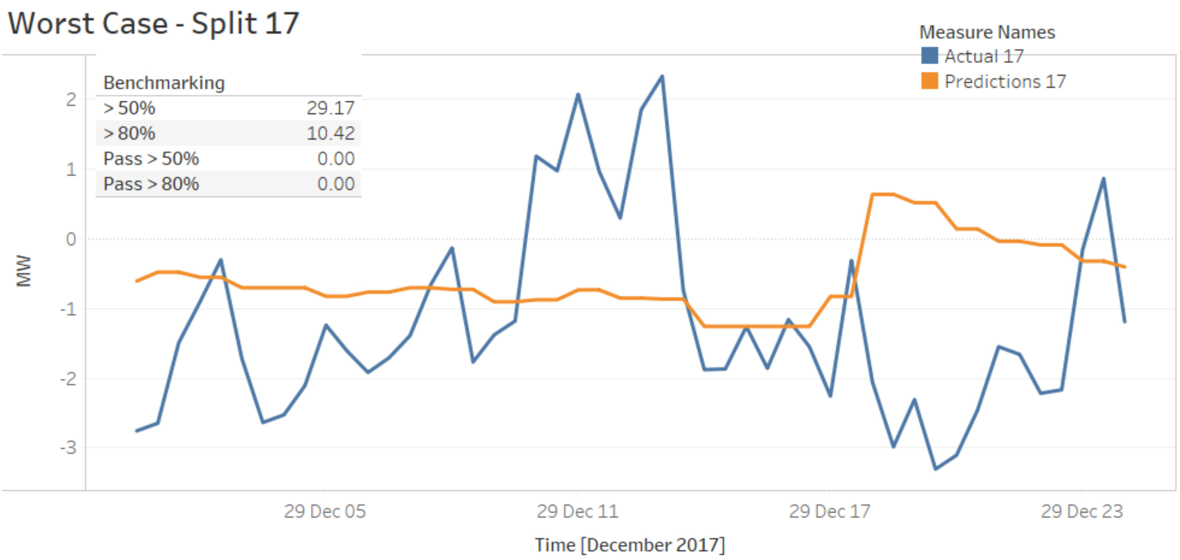


Figure 28. Primary – Kingsweston – 1 day





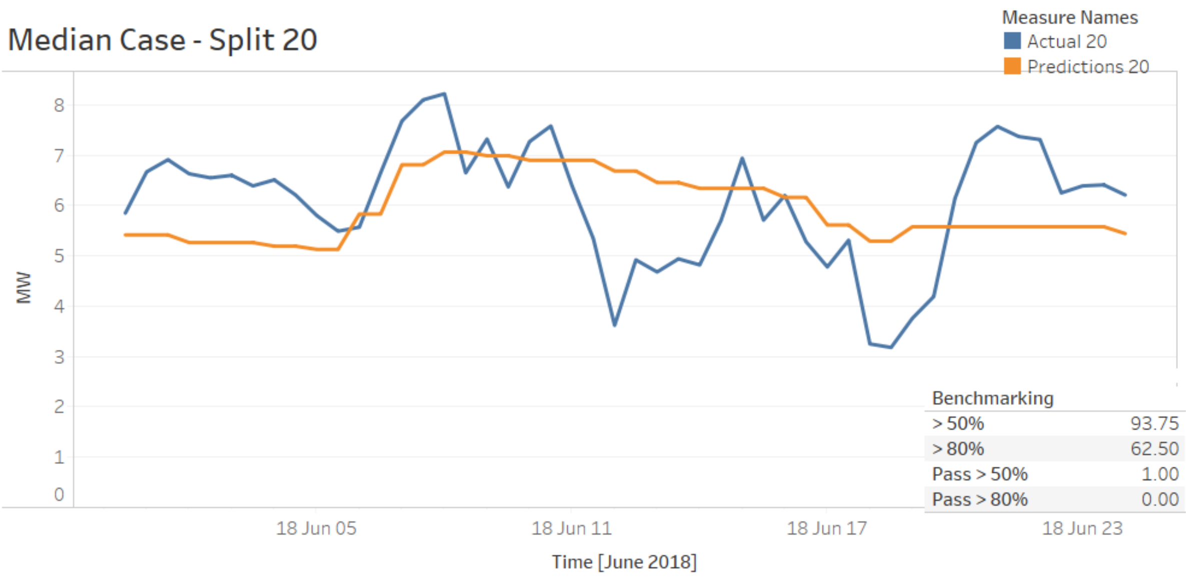
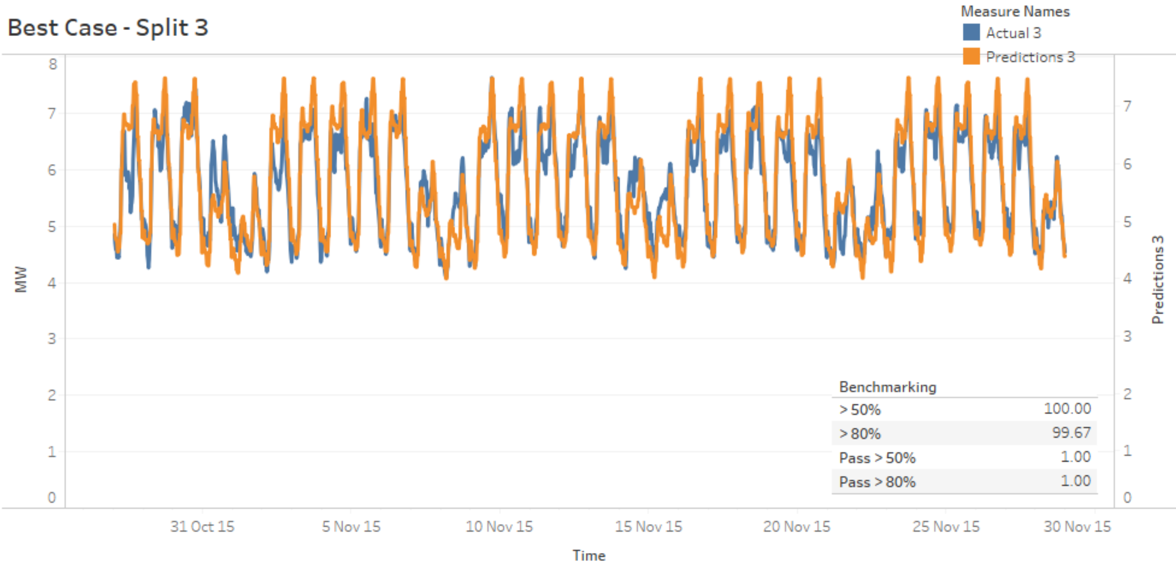
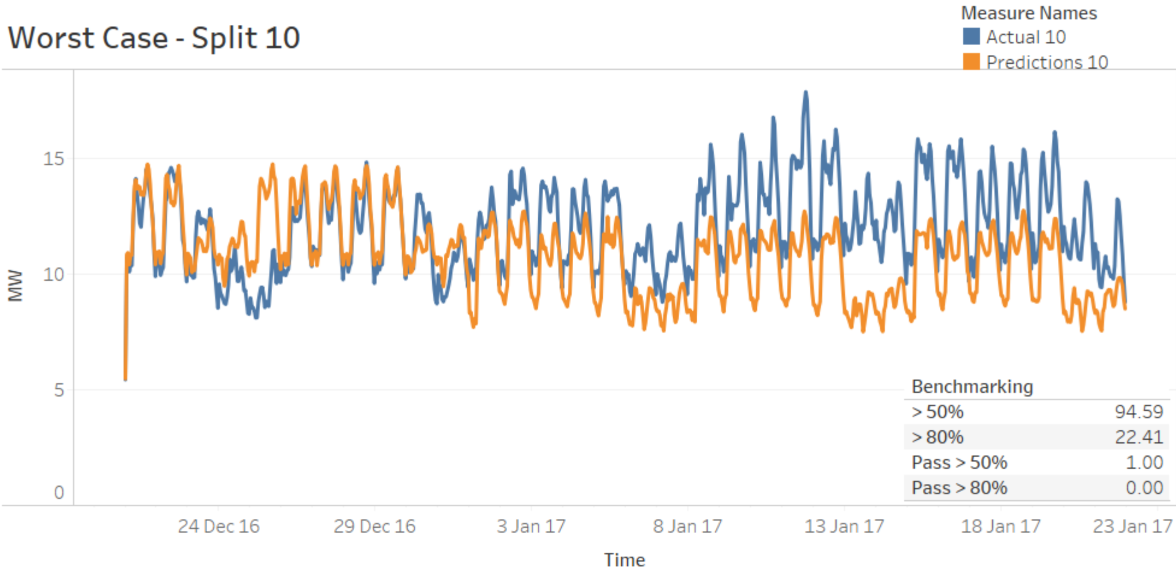


Figure 29. Primary – Evercreech – 1month





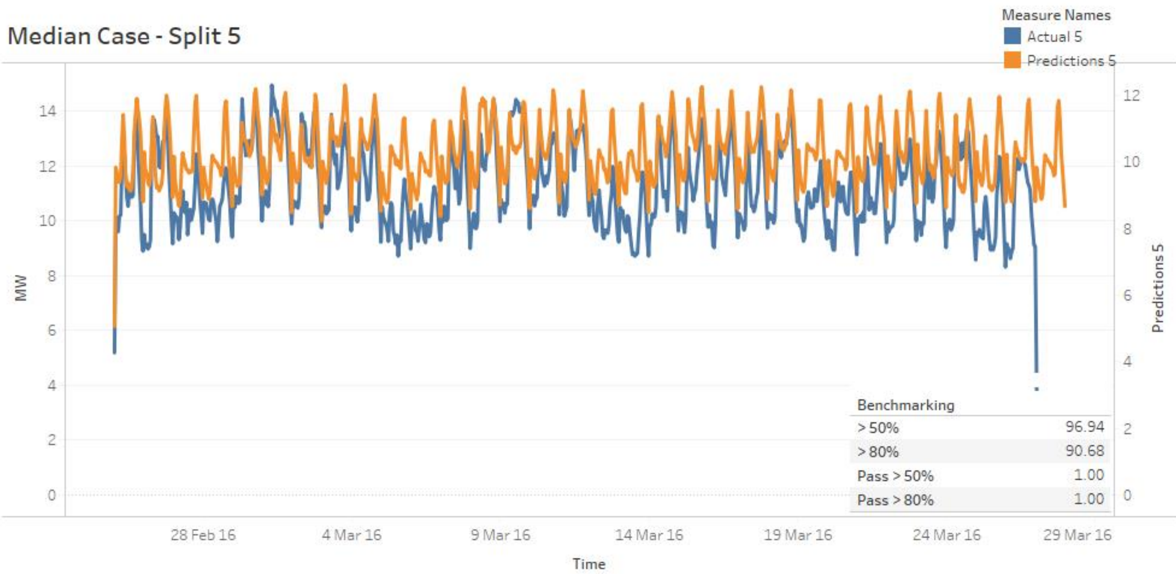


Figure 30. Primary – Evercreech – 1 day

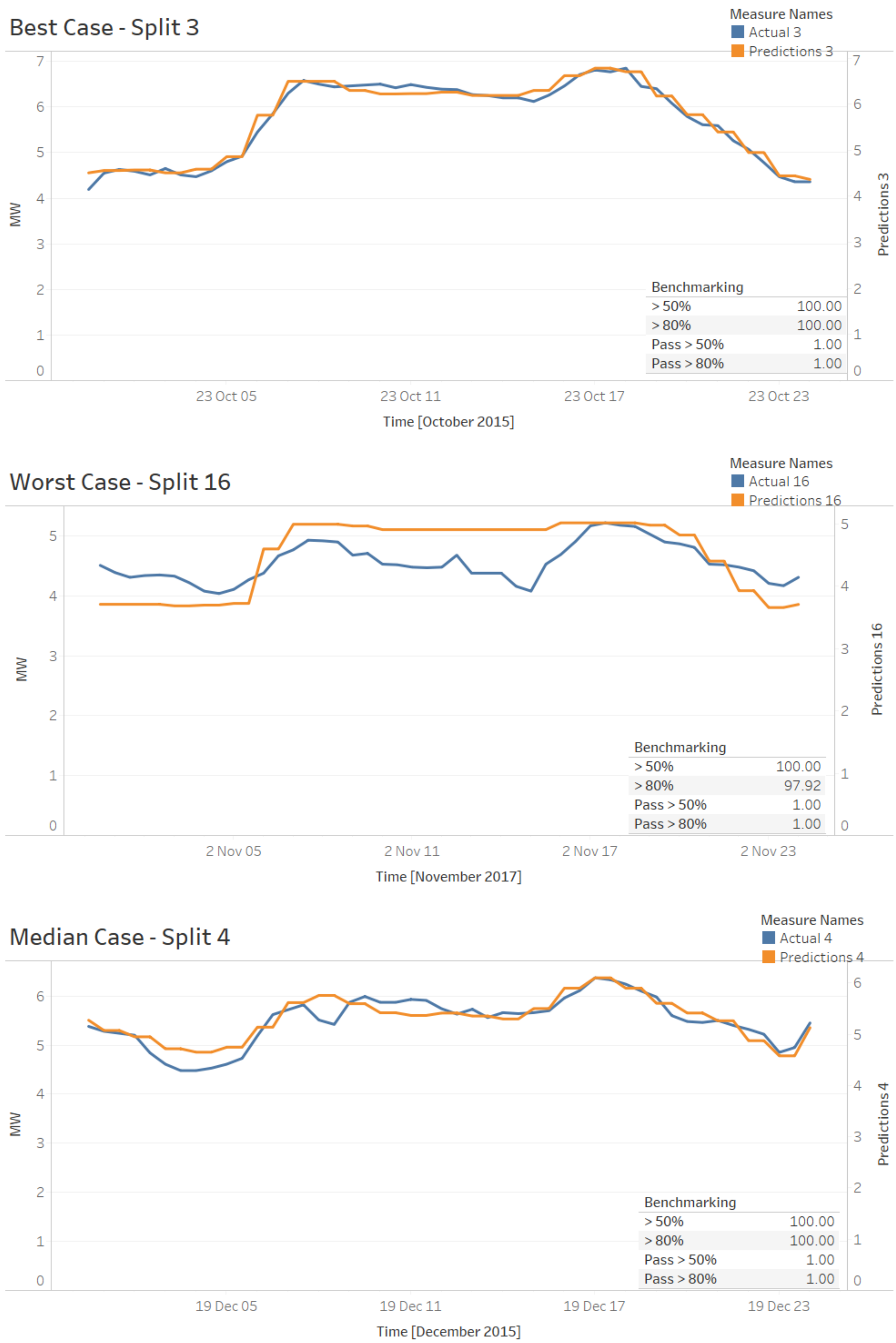
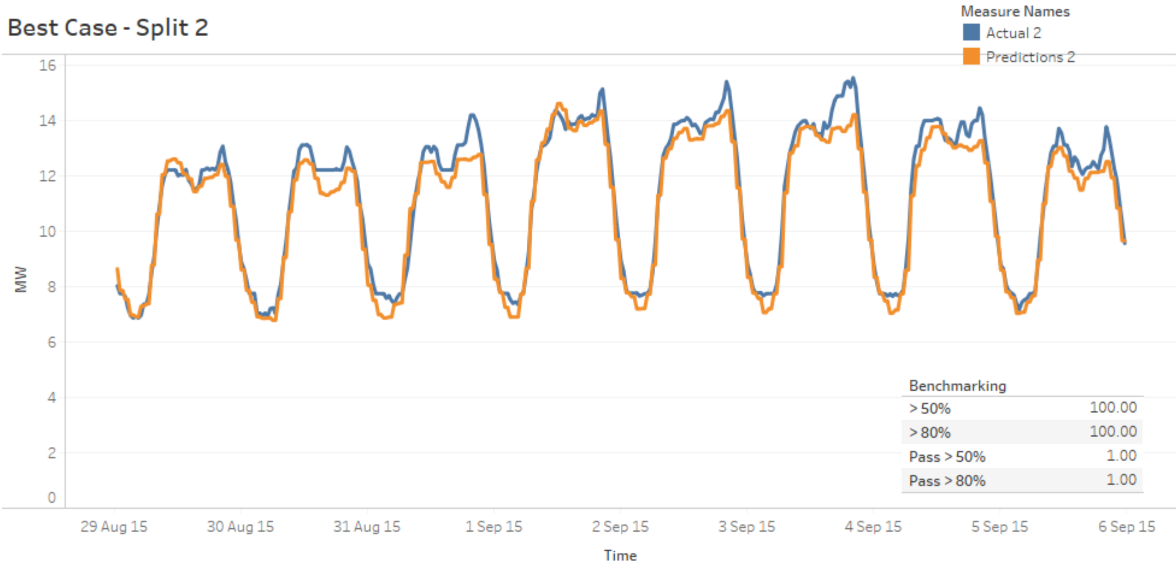
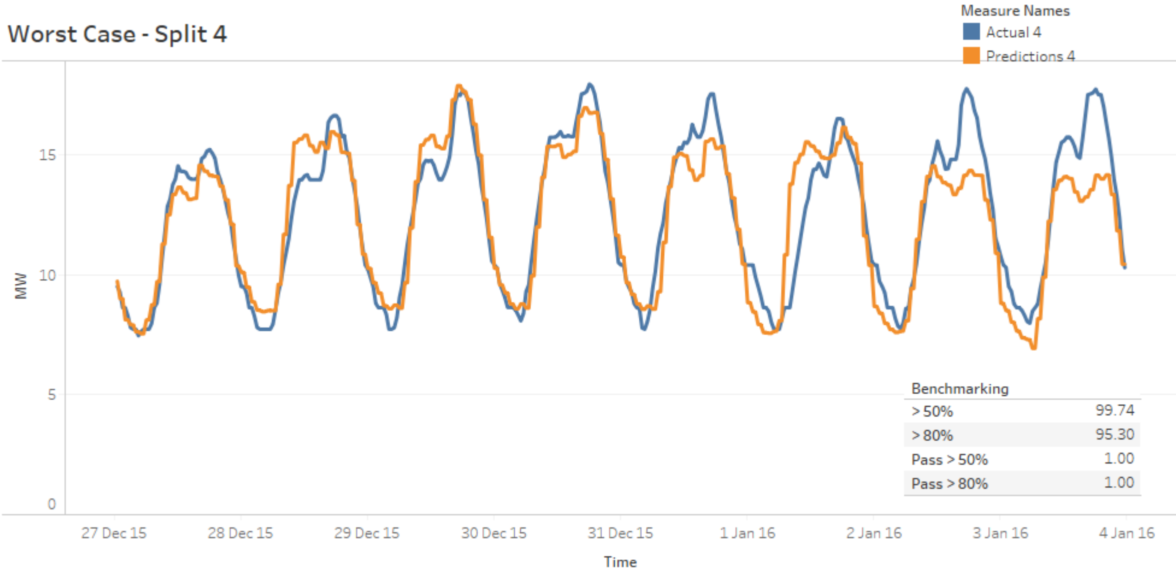


Figure 31. Primary - Cardiff East – 1 week





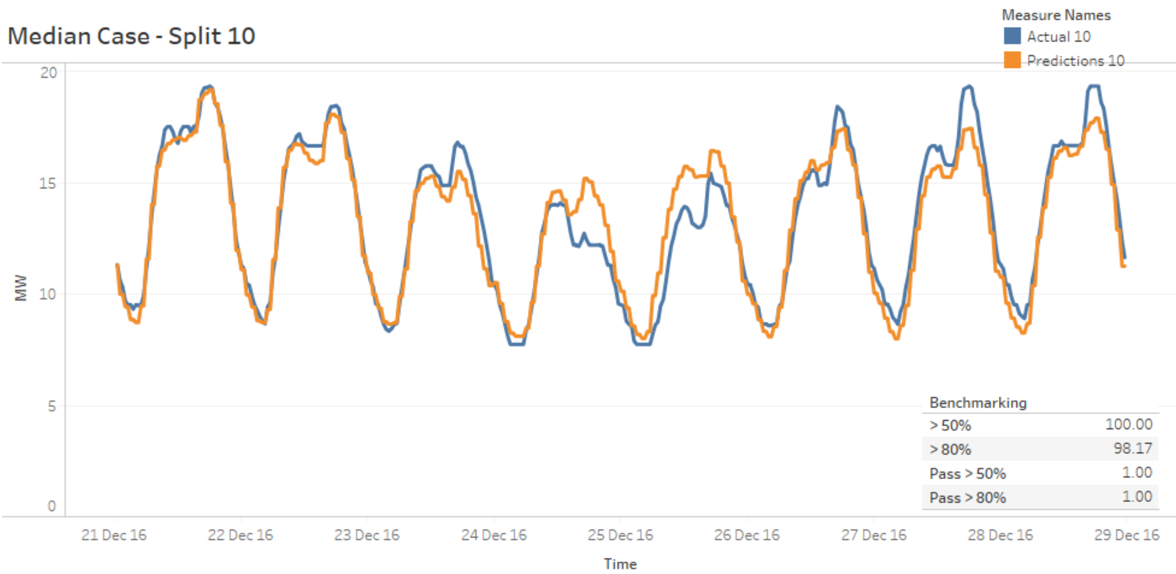
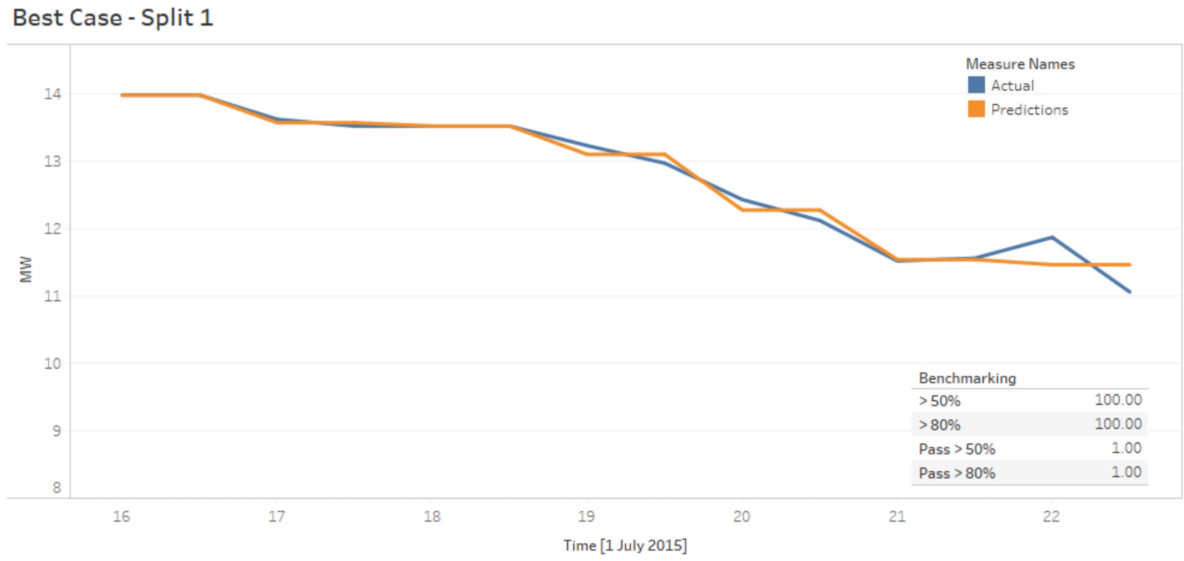
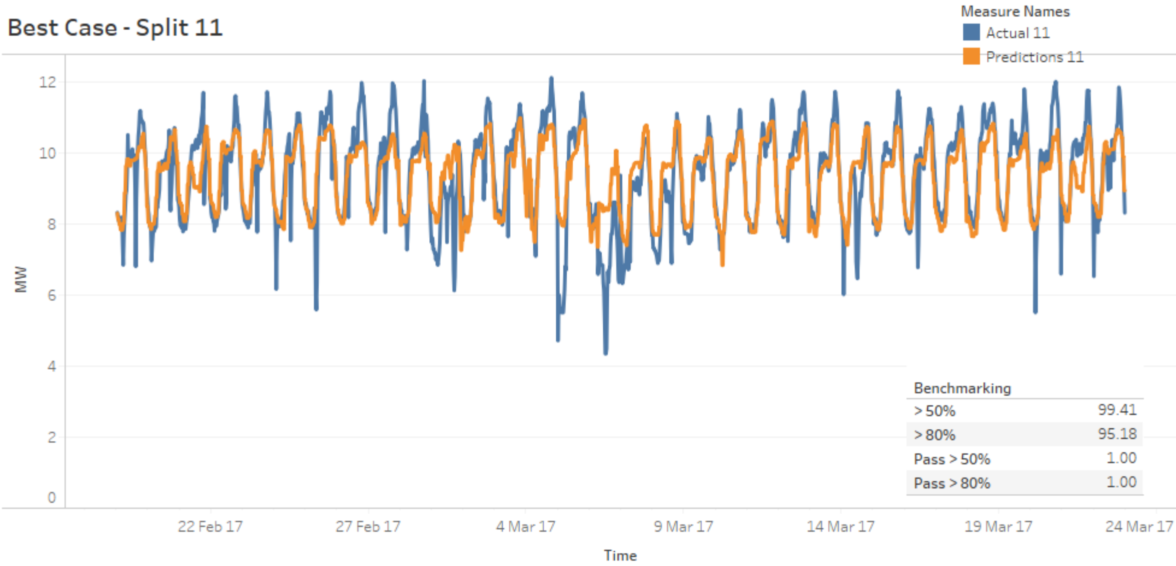


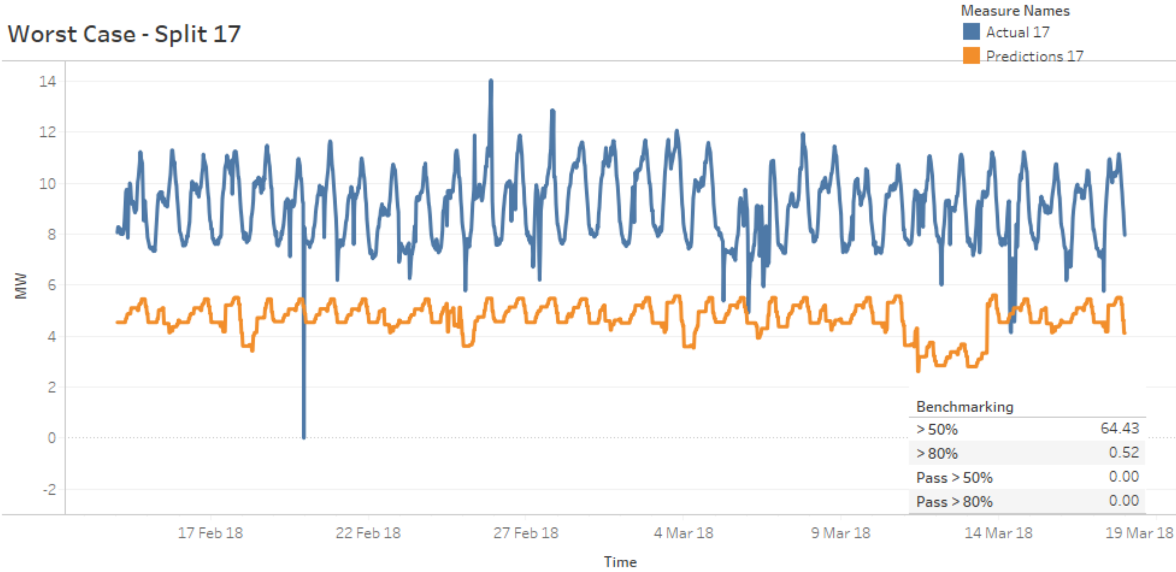
Figure 32. Primary – Cardiff East – 1 hour



Note: All cases meet 100% the benchmarking for all BSPs

Figure 33. Primary – Llynfi – 1 month





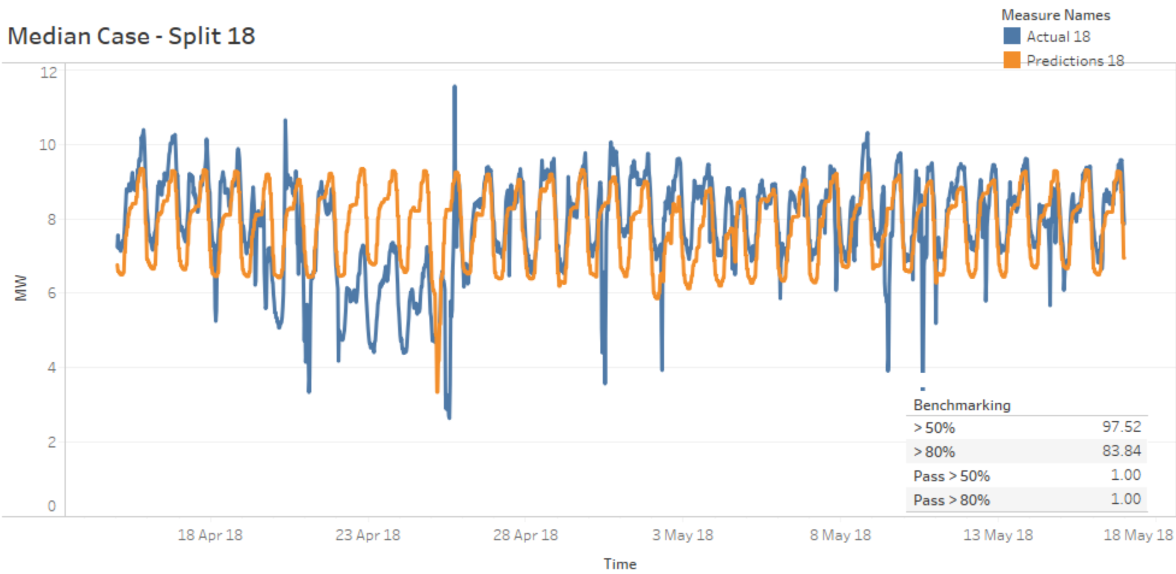
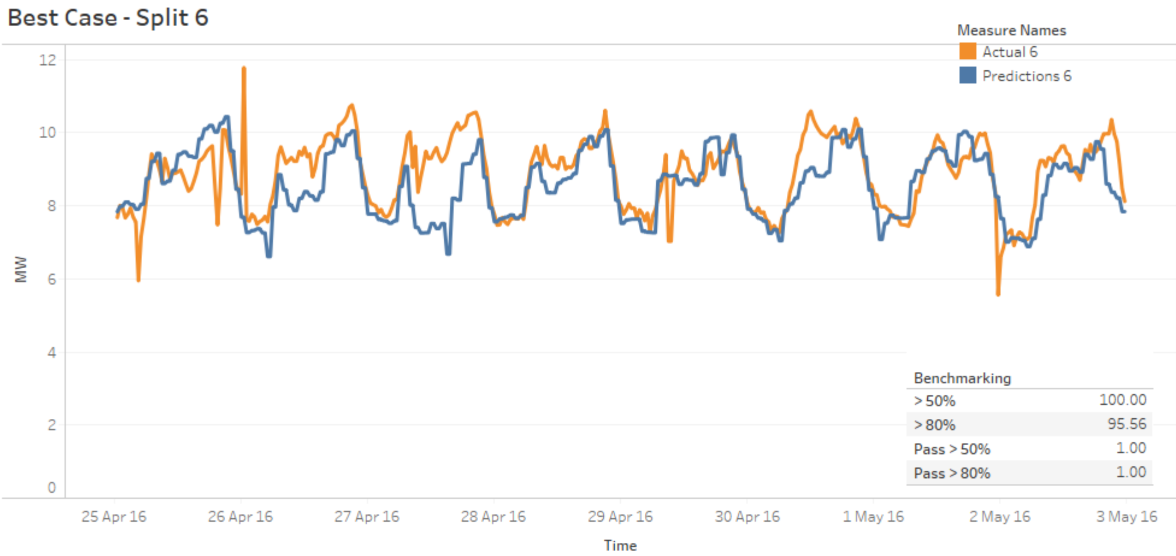
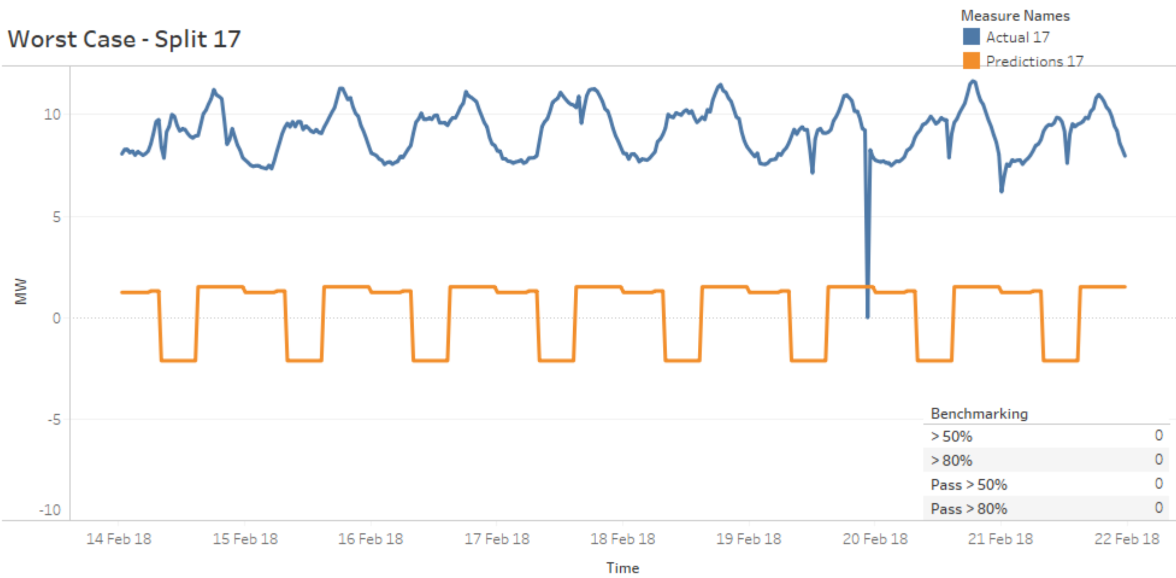


Figure 34. Primary – Llynfi – 1 week





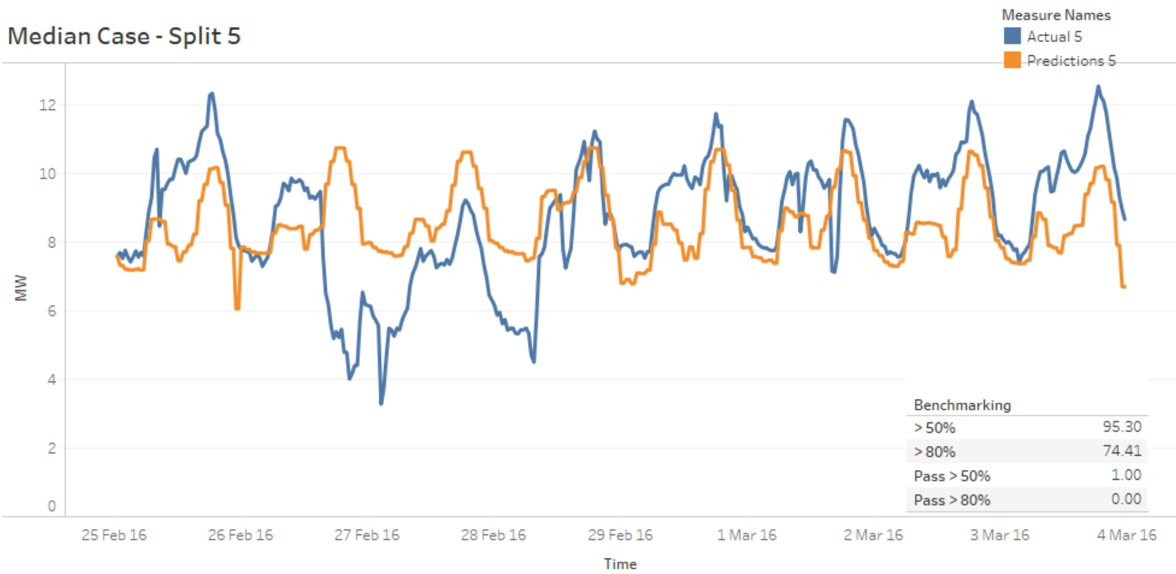
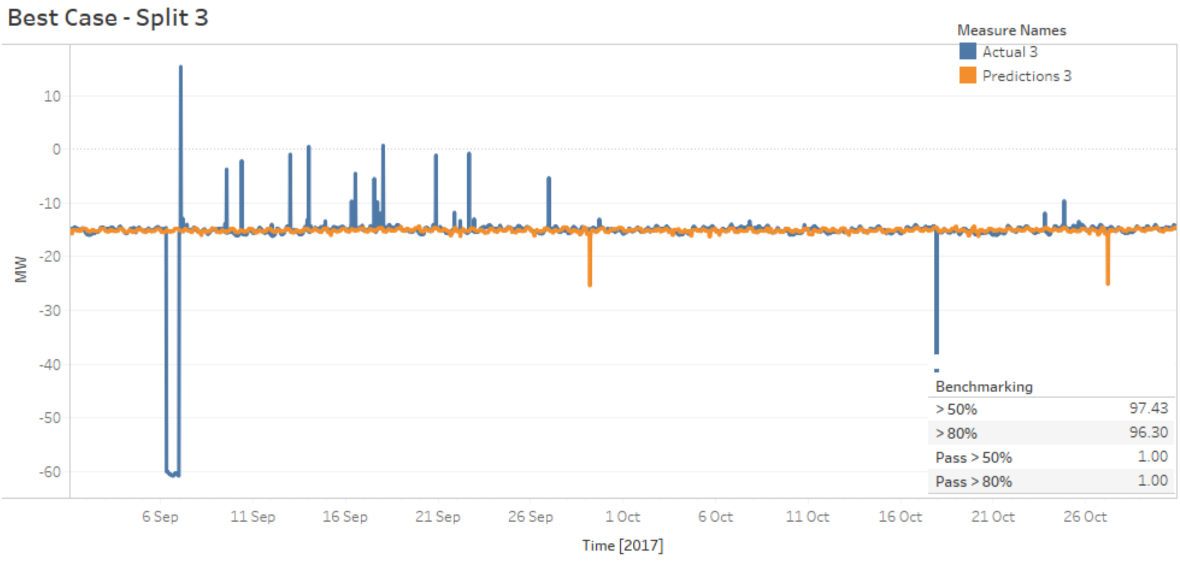
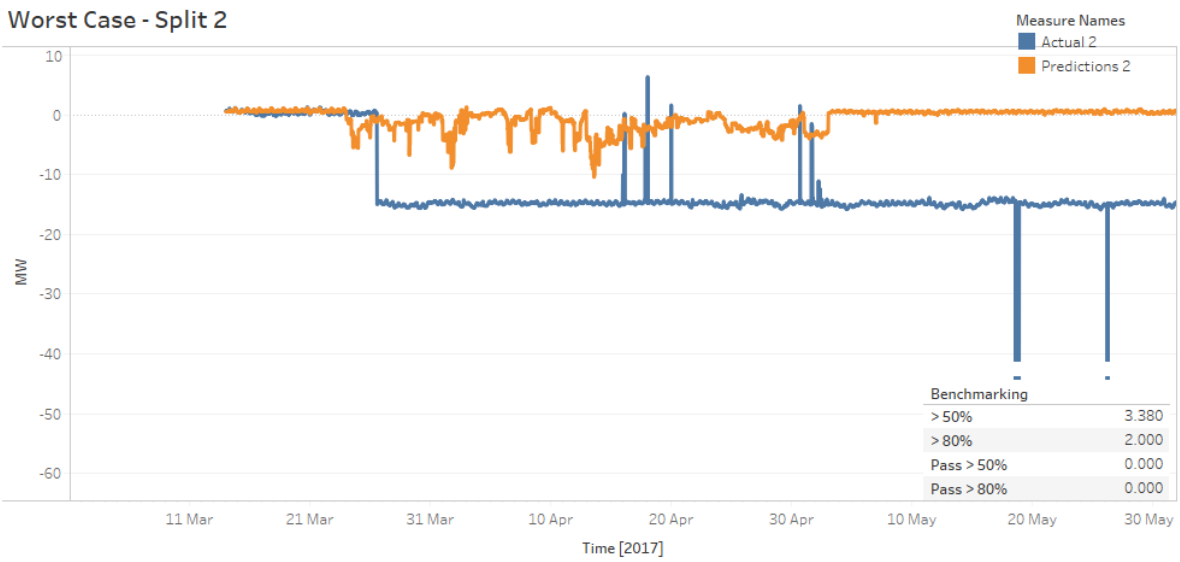


Figure 35. Primary - St Clears – 6 months





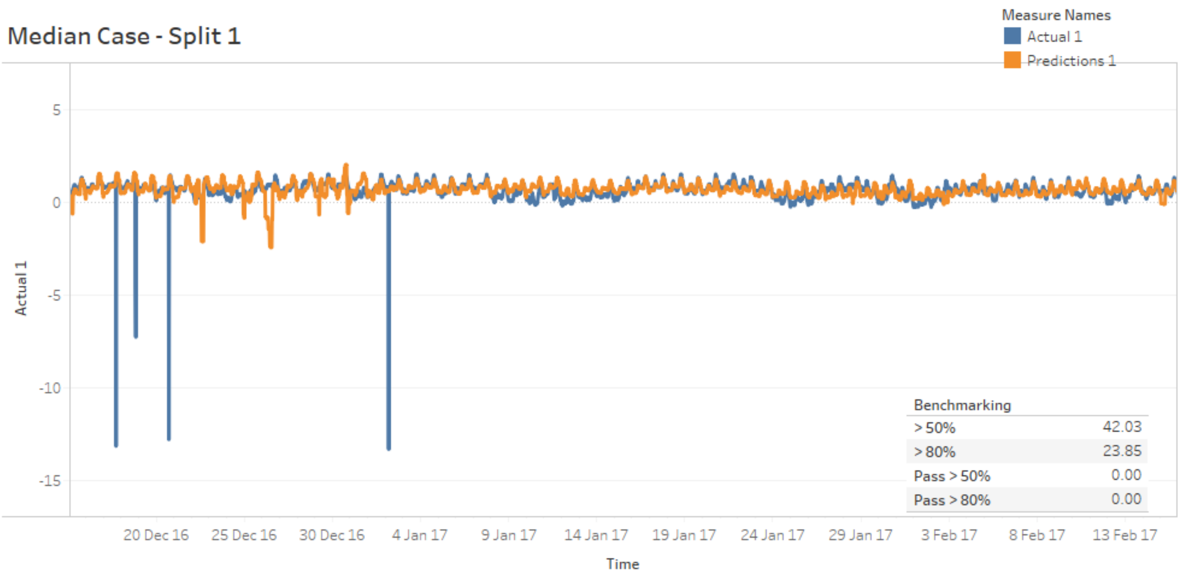
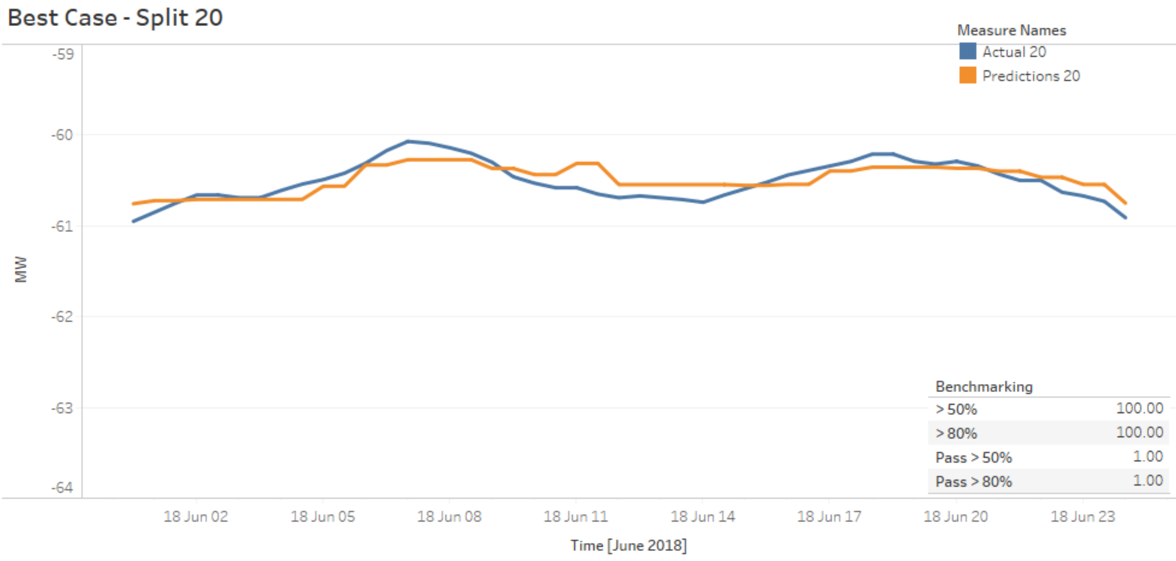
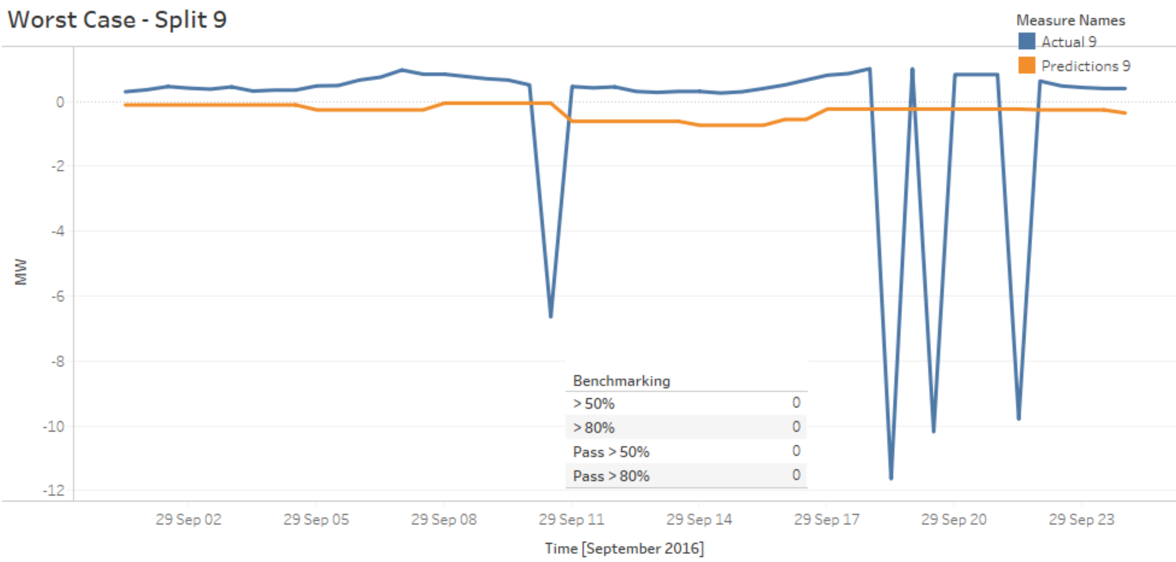
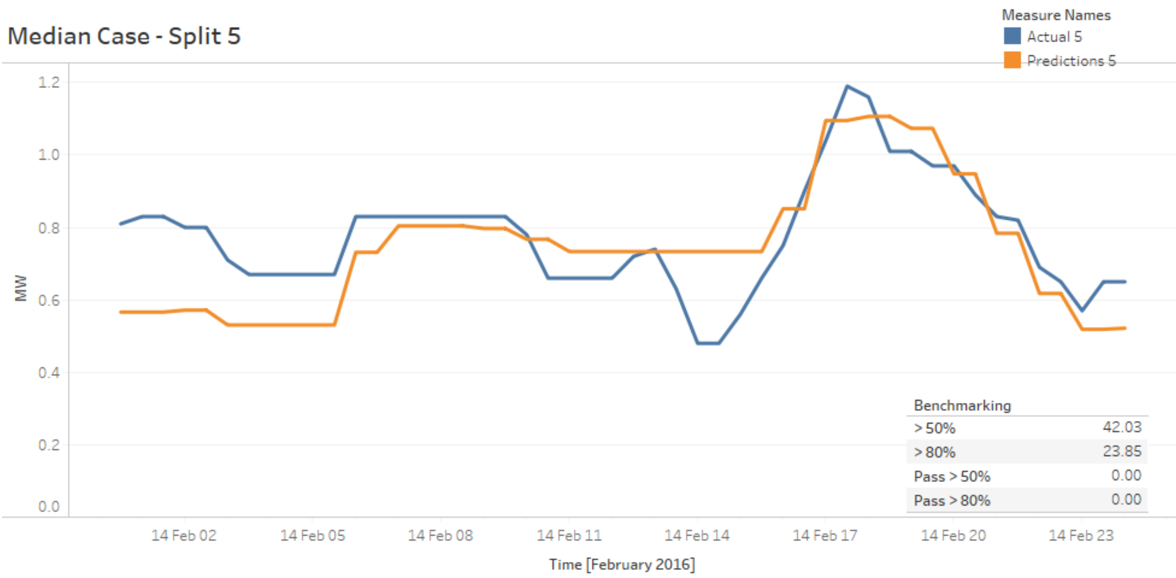


Figure 36. Primary – St Clears – 1 week

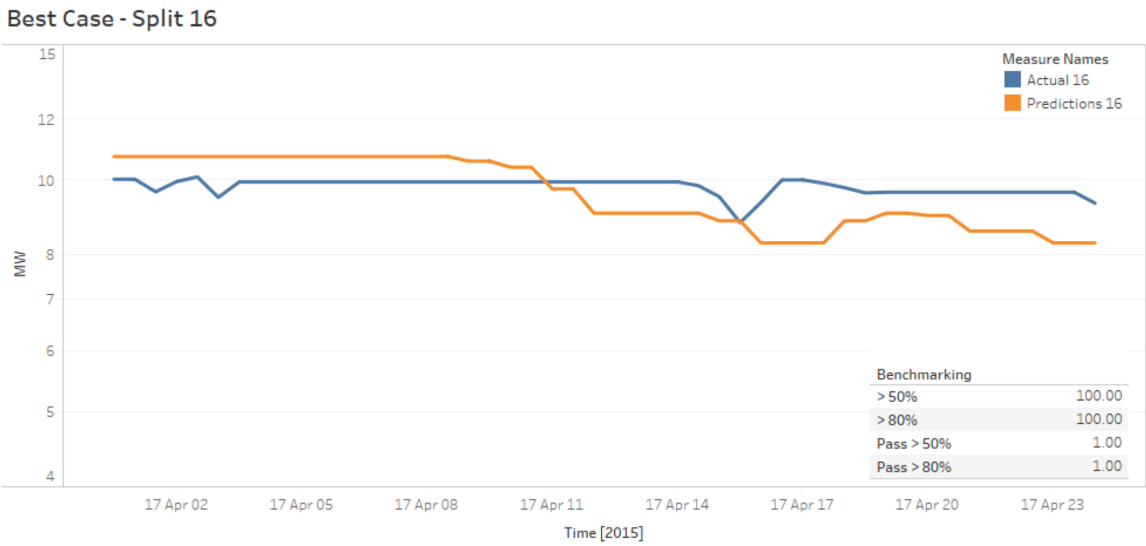


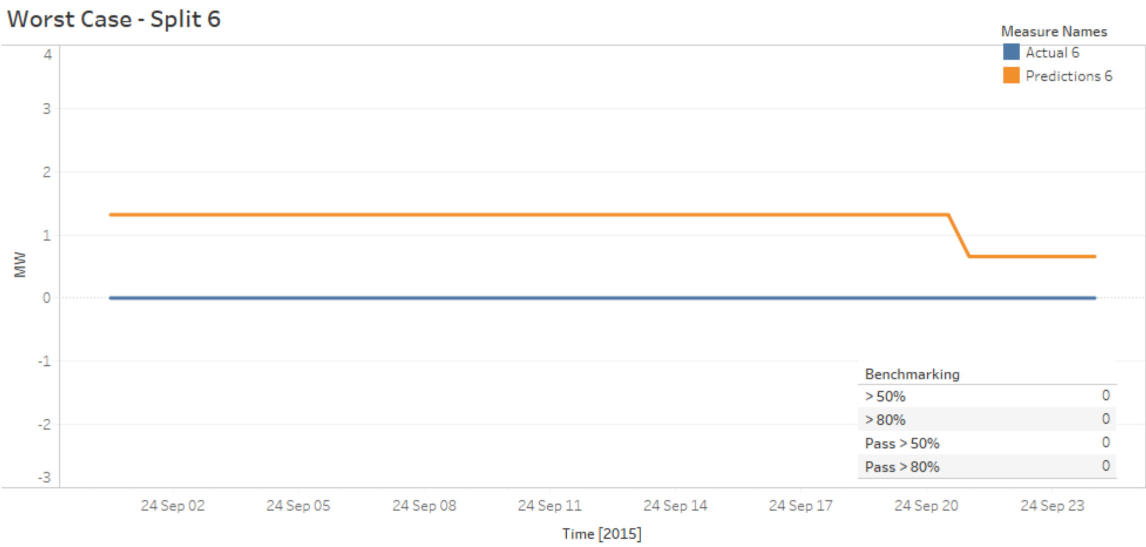




6.4. Generation Customers

Figure 37. Generation – Goonhilly Wind Farm – 1 day





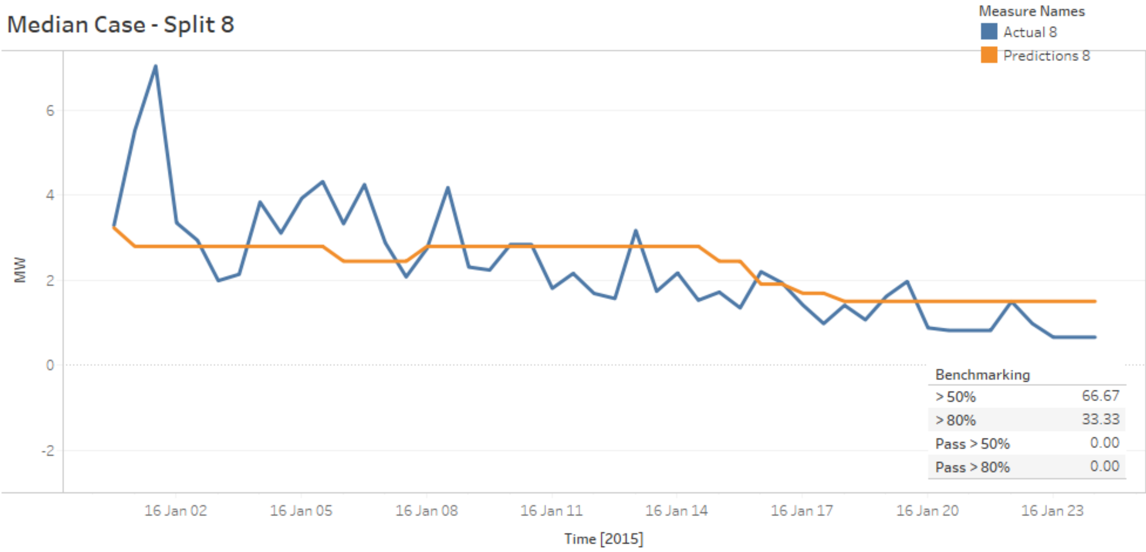


Figure 38. Generation – Rockhead Wind Farm – 1 hour

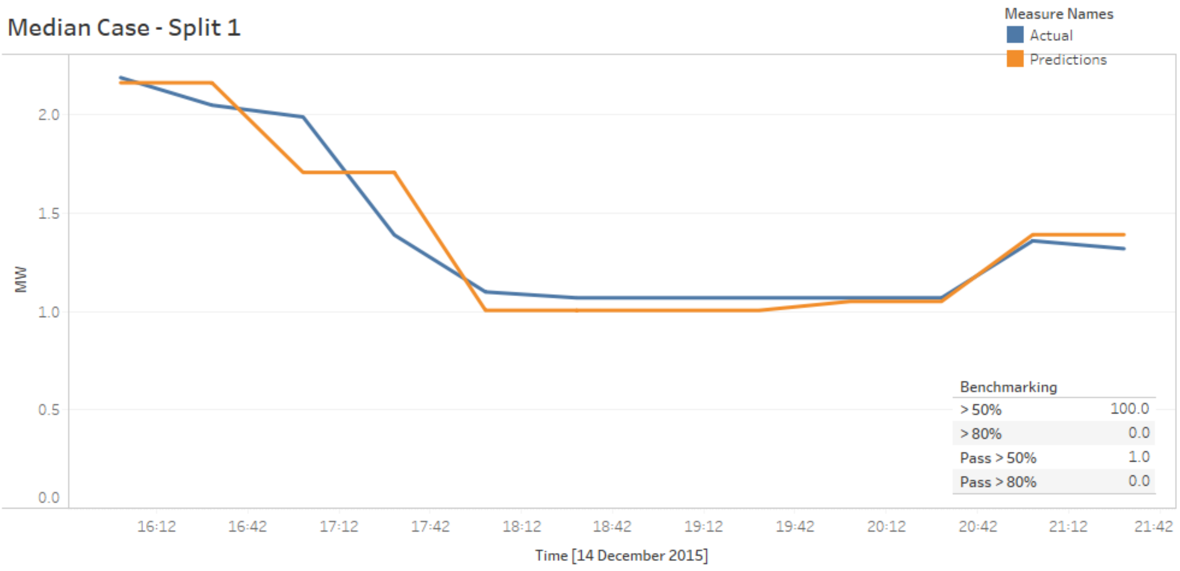
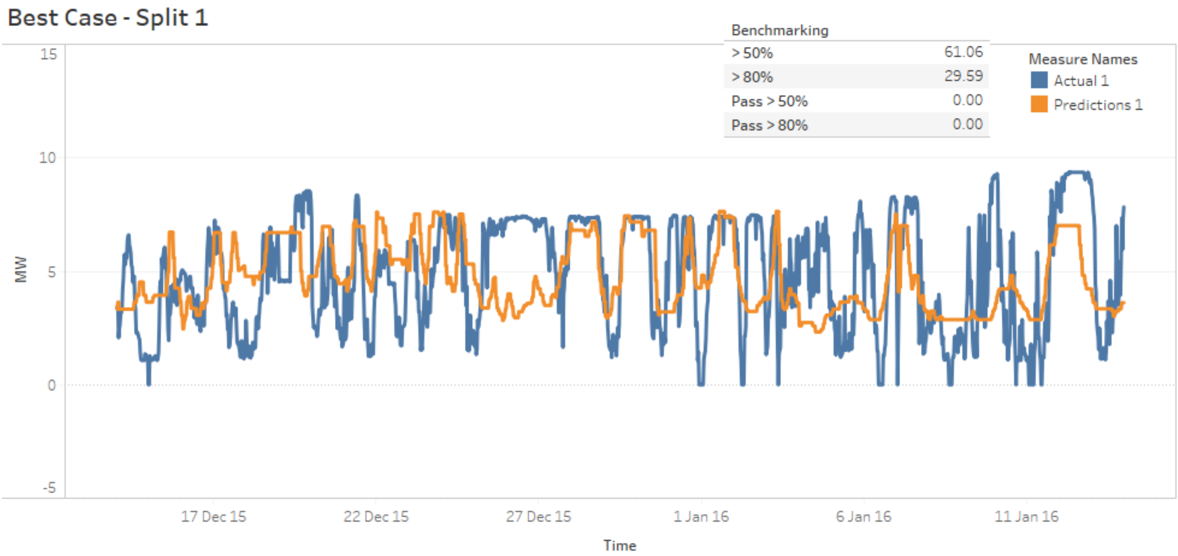
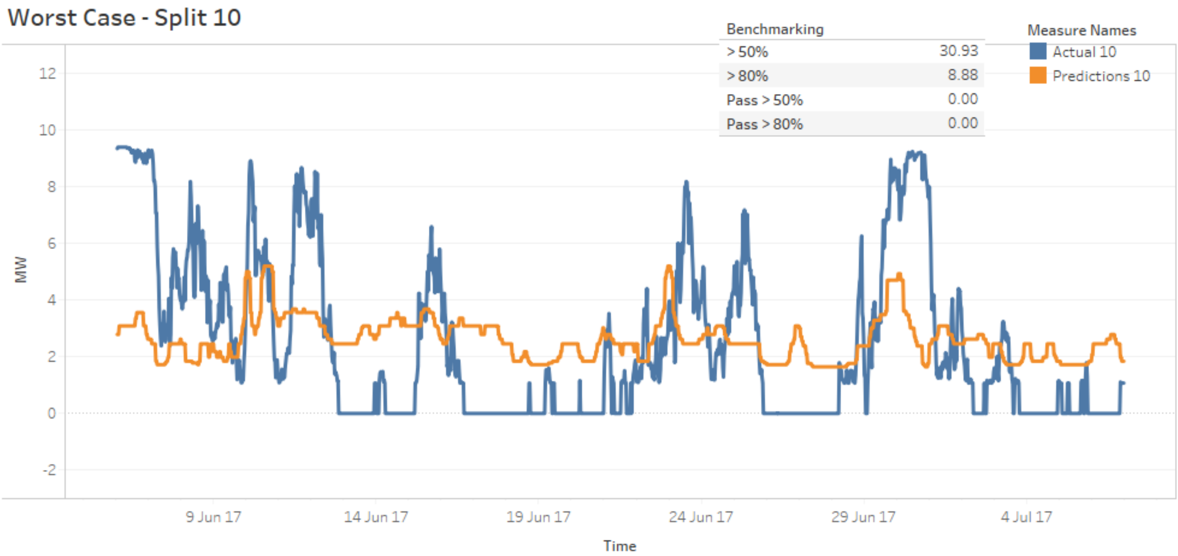
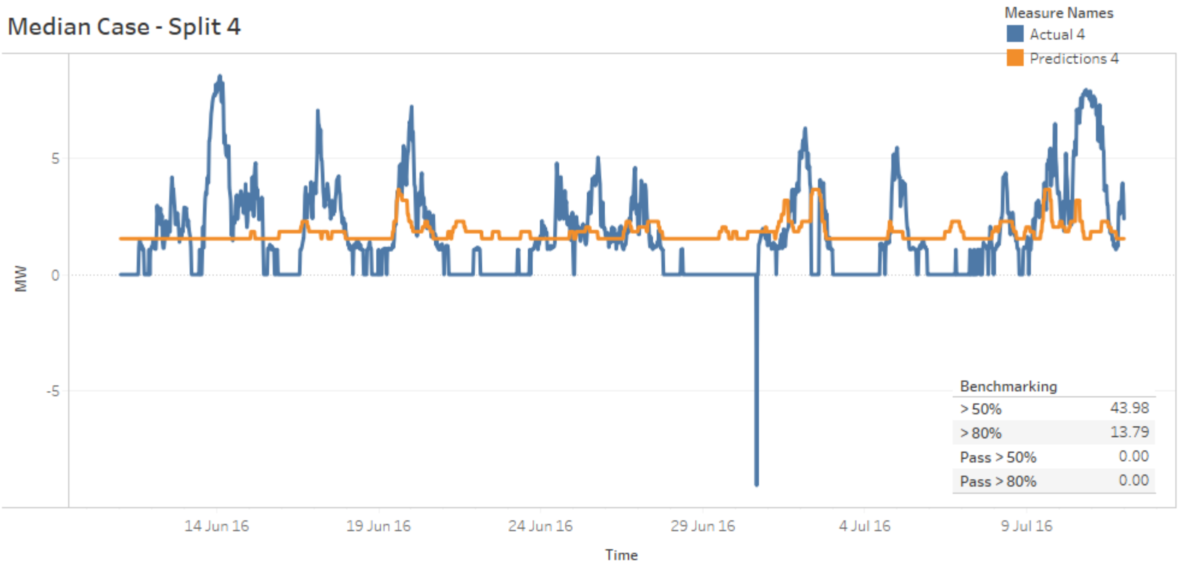


Figure 39. Generation – Rockhead Wind Farm – 1 month







6.5. Load Customers

Figure 40. Load customer – Durabill – 11*59 – 1 day

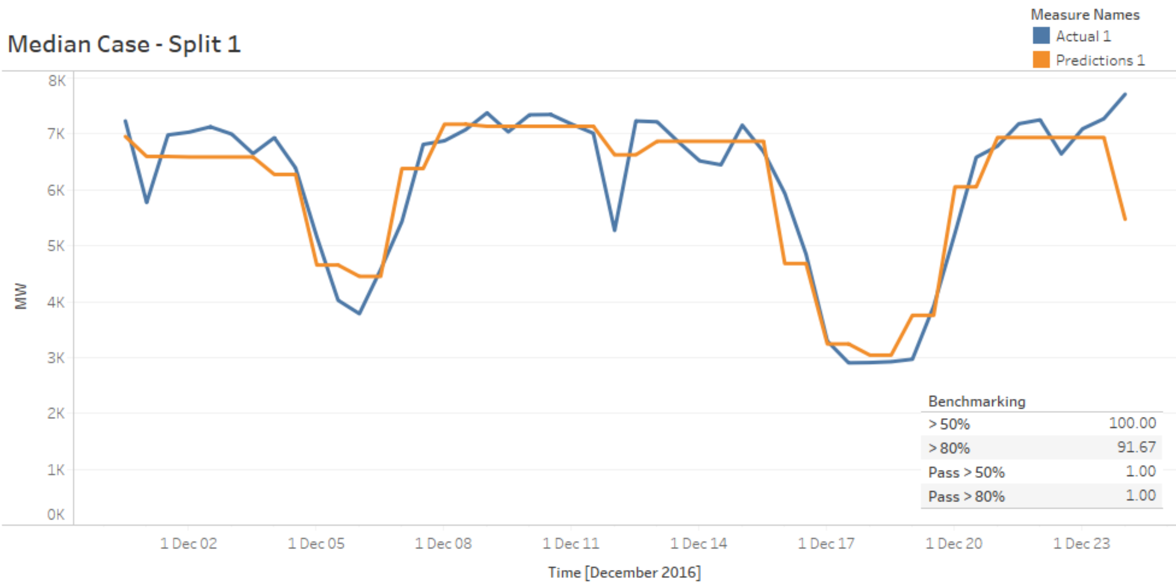


Figure 41. Load customer – Durabill – 11*59 – 1 hour

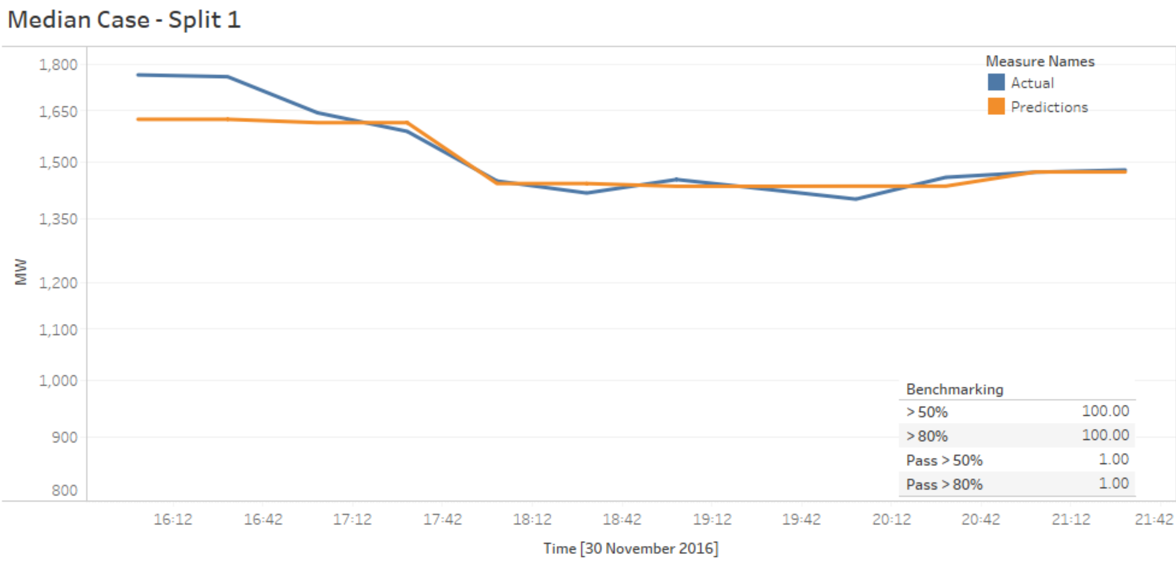


Figure 42. Load customer – Durabill – 11*58- 1 day

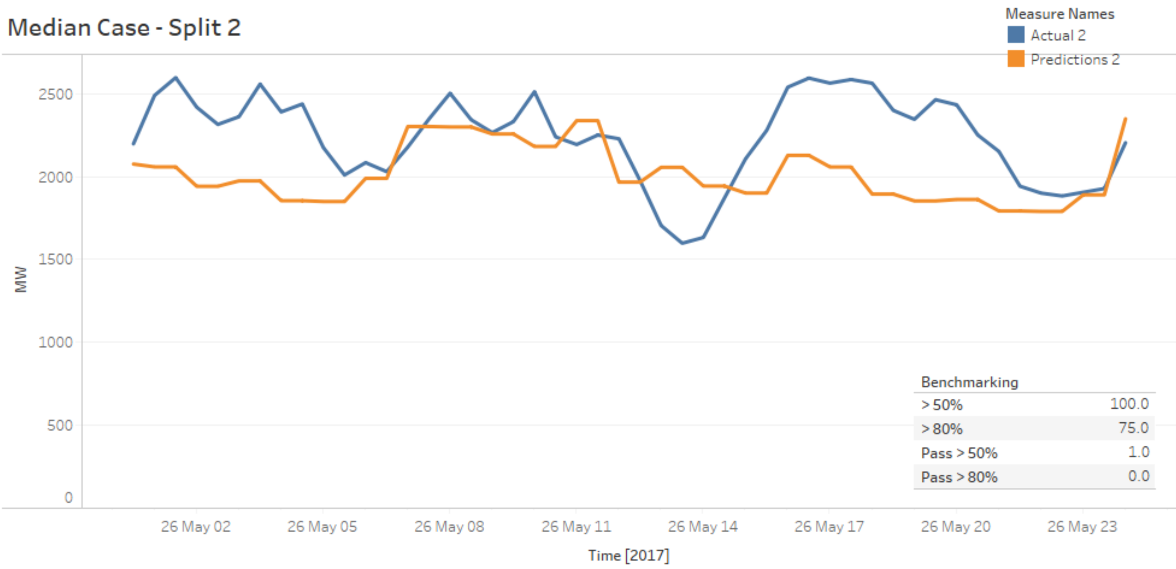


Figure 43. Load customer – Durabill – 11*94 – 1 hour

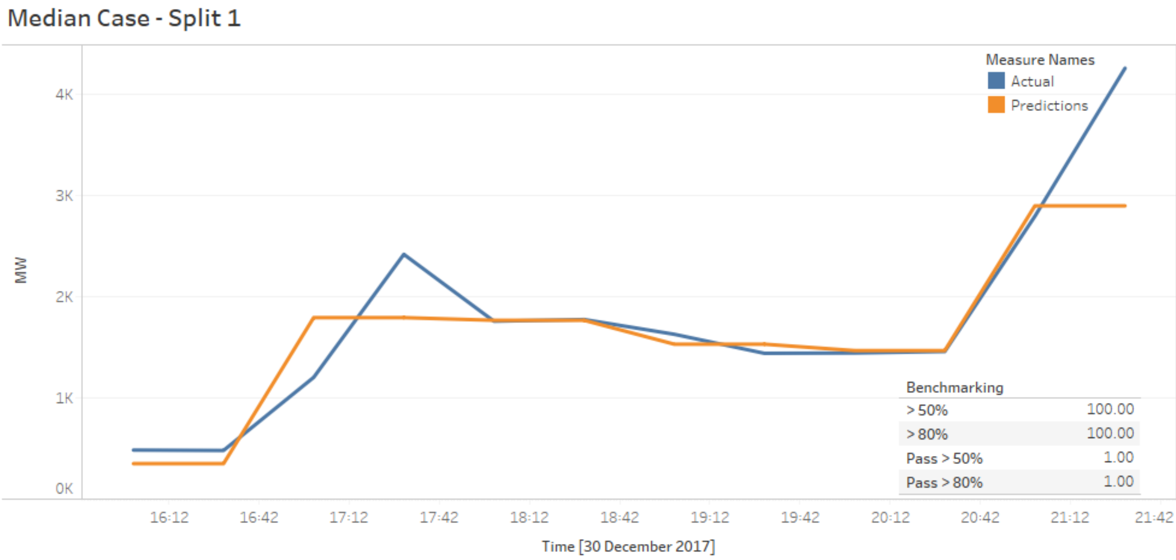


Figure 44. Load customer – Durabill – 14*00 – 1 day

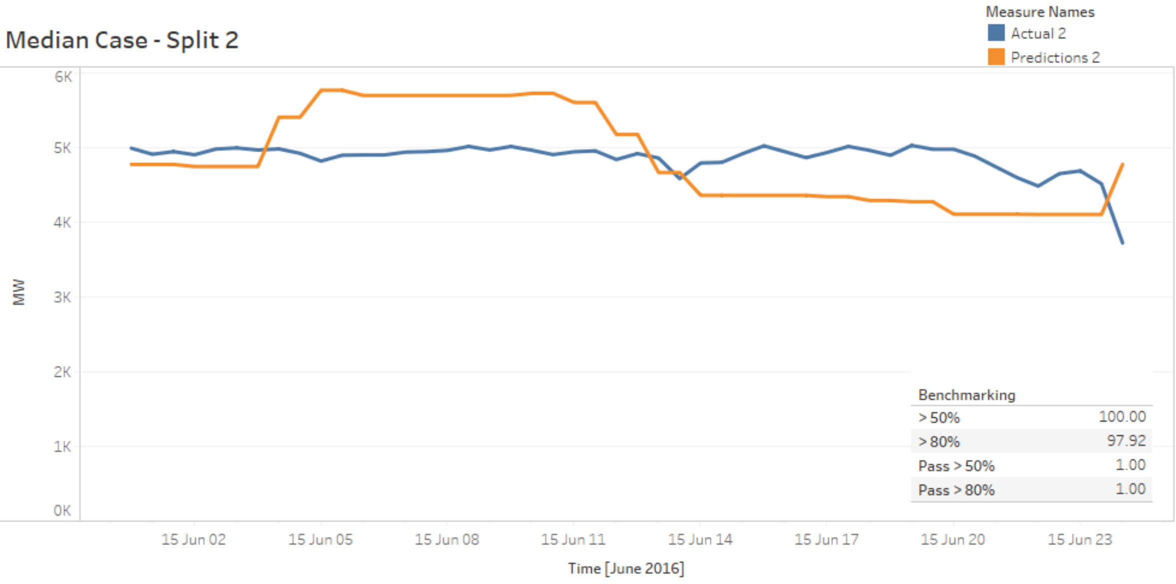


Figure 45. Load customer – Durabill – 14*08 – 1 hour

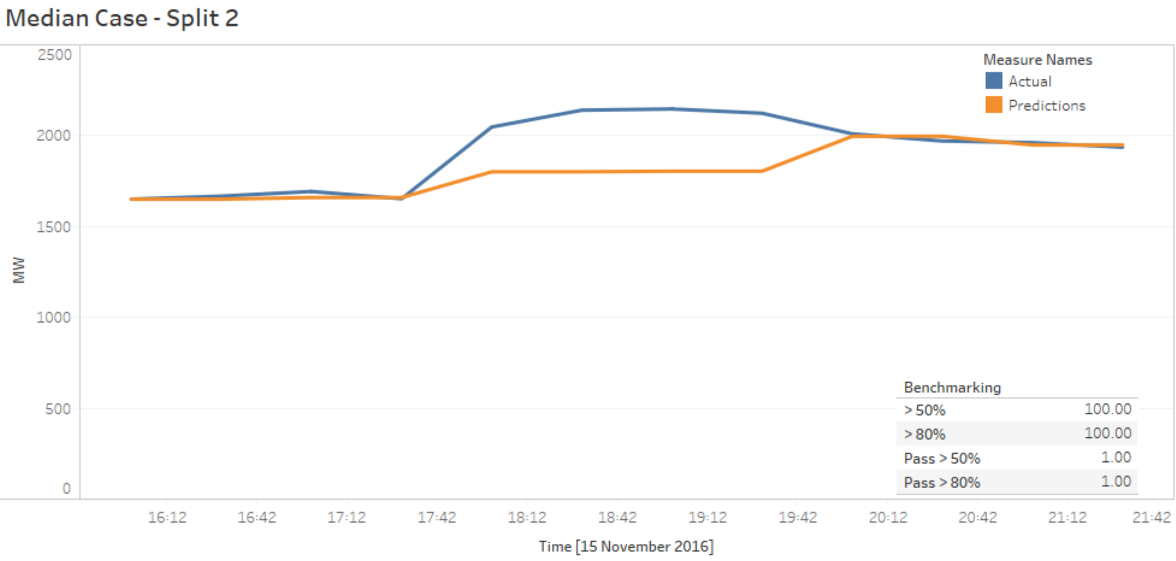


Figure 46. Load customer – Durabill – 14*05 – 1 day

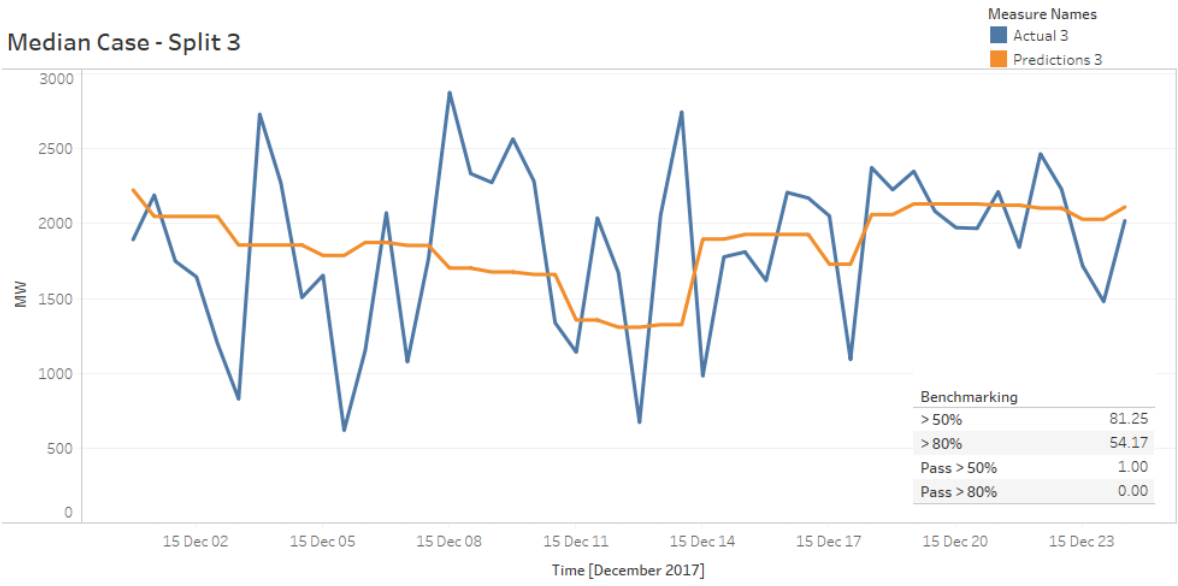


Figure 47. Load customer – SCADA – Wymeswold – 6 months

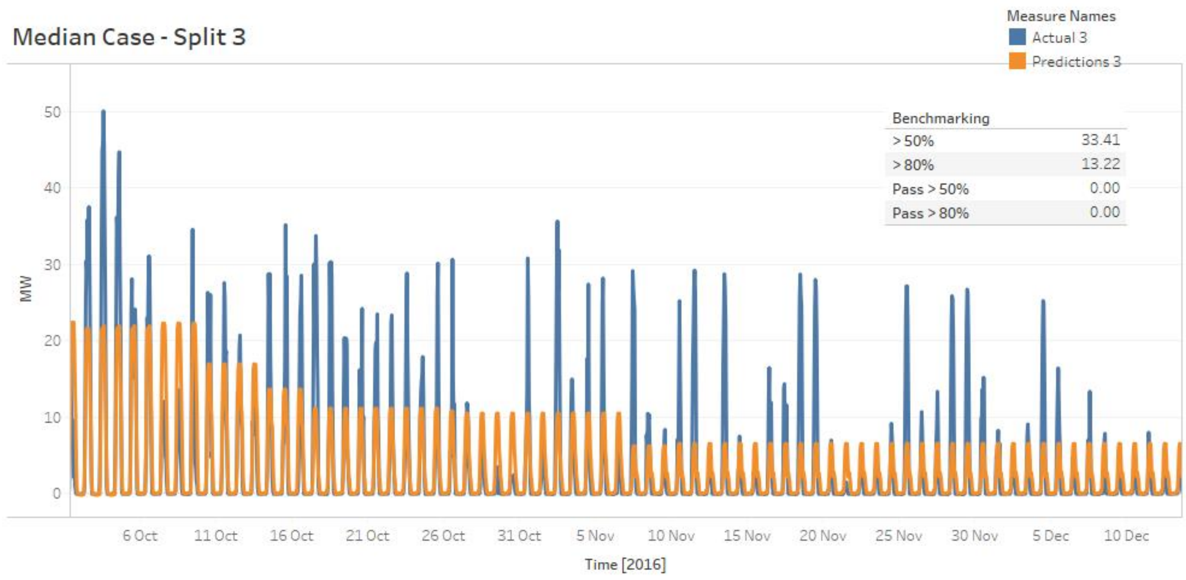


Figure 48. Load customer – SCADA – Wymeswold – 1 week

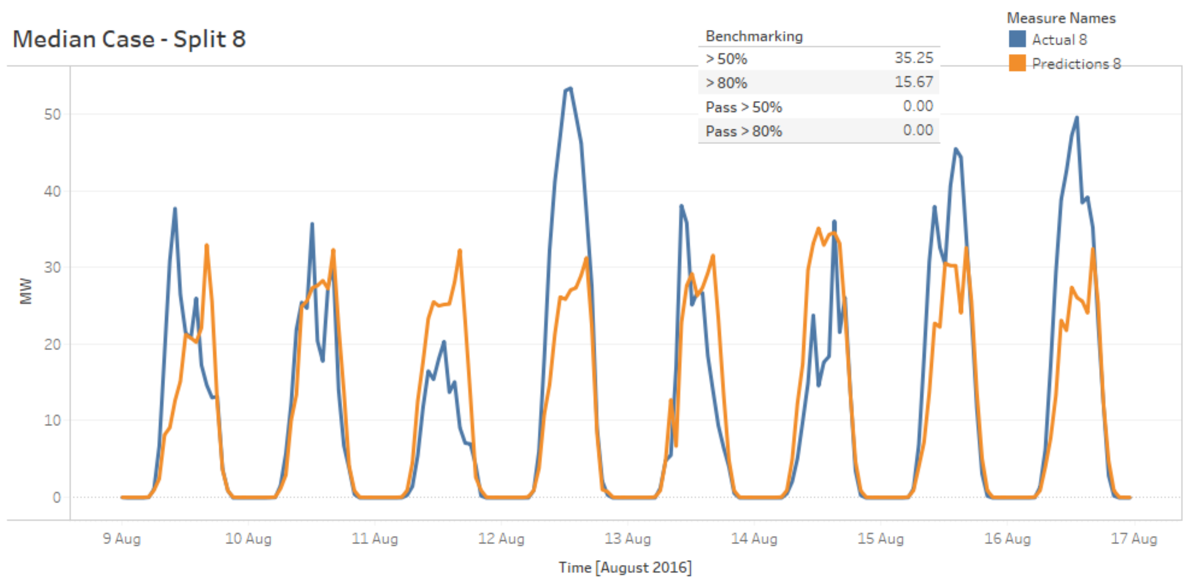


Figure 49. Load customer – SCADA – Jaguar Land Rover – 1 month

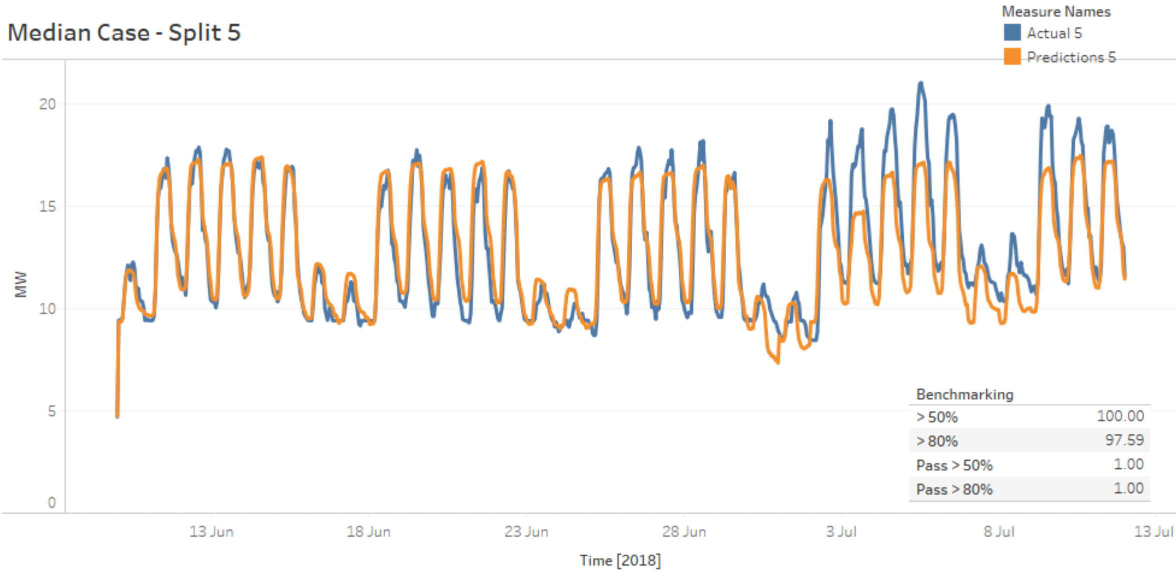
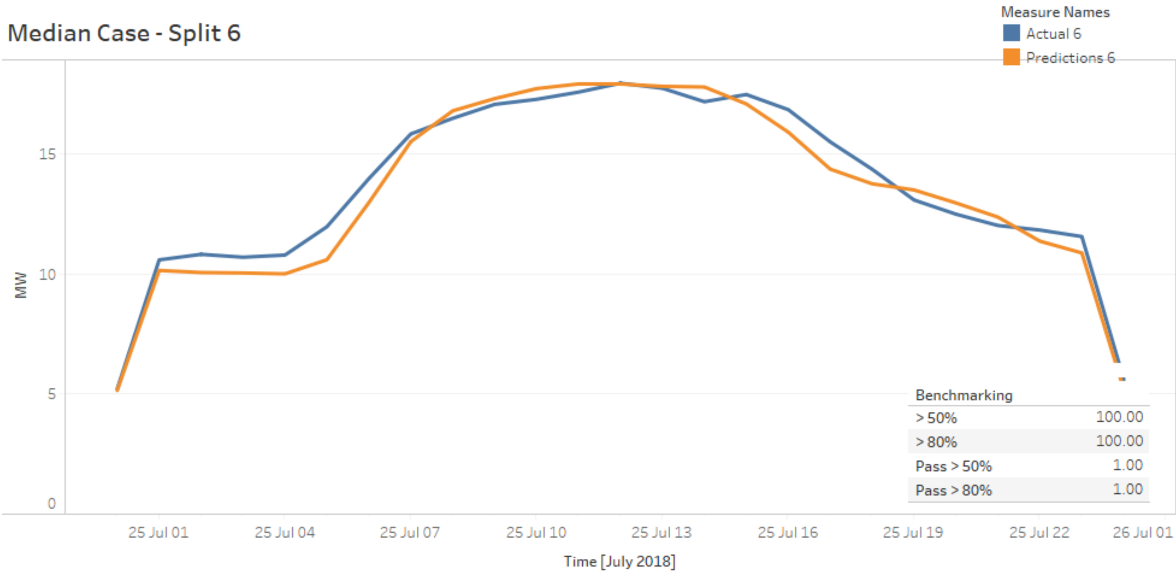


Figure 50. Load customer – SCADA – Jaguar Land Rover – 1 day



7. Conclusions and Recommendations

7.1 Replication of SGS toolchain and results

The open source toolchain proposed by SGS to generate forecasts has been replicated by Capita DA. Further, when applying the methodology proposed by SGS on the same use locations, Capita DA has obtained very similar results to those reported by SGS - suggesting that the replication had successfully been achieved. In cases where SGS achieved high accuracy (e.g. UC2 and UC3), Capita DA has achieved equivalent levels of accuracy, with small differences observed due to differences in the sampling methodology and the inherent randomness of the XGBoost model. Where results reported by SGS were less accurate (e.g. UC1, UC4, UC5, UC6), Capita DA has achieved comparable levels of accuracy.

Where model results were poor, this was often the result of data quality issues reflecting that the models were built using uncleansed data.

Different locations exhibit different levels of randomness in their behaviour. The model worked best on locations where a consistent pattern is observed and no data quality issues were present. In these cases the model achieved consistently high accuracy even for the longest time horizons. In cases where data quality was sufficient but behaviour is more random, the model performance varied substantially. Capita DA proposes performing a model optimisation procedure by experimenting with different features and training times until the best performing combination is found.

A default set of hyperparameters for each forecast site family should be provided, though the ability to tailor features for each site should be retained. Once a wider set of results has been achieved it may be possible to set up rules for acceptance e.g. forecasts where day ahead, week ahead forecasts meet 80% criteria and month ahead meets 50% criteria could be considered adequate. These rules would filter out those models that need further tailoring from those that can work acceptably using the default feature set or other default variables.

Similarly, more experience is needed at creating a wider set of forecasts before shortcuts to find the optimal the training period for different time horizons can be proposed. It is not clear whether there should be a default value per site family, for example.

While useful for development, the open source toolchain suggested by SGS is not the only mechanism by which XGBoost based forecasts can be made. For further development of the EFFS project, WPD and other DSOs can further develop the existing functionality in a way which better supports automation.

7.2 High-level conclusions

The validation testing exercise allowed some high-level conclusions to be drawn in terms of understanding where the model performs well and where further attention is needed.

1. Data quality is key to achieving reliable forecasts. In several test cases the input data contained outliers, zero readings and error readings that hamper quality of forecasts (affecting the training data, test data, or both). Capita DA's recommendation is to introduce a data audit and data cleansing process prior to forecasting, and to undertake this exercise as a distinct activity before the transition to BAU. Given that the EFFS trial will require good quality forecasts, data quality for the trials areas should be investigated as soon as possible to allow time to cleanse the data;
2. In general, shorter time horizons work better
3. GSPs were tested by modelling each transformer separately. The higher level of network aggregation at GSP level makes their behaviour more stochastic compared to e.g. BSPs and therefore harder to predict. The set acceptance criteria can only be reached for hour ahead forecasts on average, but not in all simulations;
4. Certain primaries and BSPs exhibit predictable load profiles and can be forecast accurately even for long-term time horizons, while others exhibit variations in the load profile that have not been explained by the model and features used, leading to inconsistency in the forecast accuracy. In these cases, the expertise of DNOs in understanding the underlying behaviour and extracting the features that might explain these variations would be a sensible next step.
5. The acceptance criteria have been met on average by three out of the four BSPs tested for week ahead and day ahead time horizons, and by all four BSPs for the hour ahead time horizon;
6. The acceptance criteria have been met on average by three of the six Primaries tested for all time horizons. Two primaries exhibited less predictable behaviour and could meet acceptance criteria for shorter time horizons, while the sixth was marred by data quality issues;
7. Two wind farms were tested, with wind data included in the feature set. It was noted that MW output is unpredictable and forecasts did not meet the acceptance criteria;
8. For load customers, two data sets were provided – SCADA data and Durabill data. It was found that Durabill data usually does not contain continuous readings required for training and testing and it was tested for day ahead and hour ahead only. On average the Durabill forecasts met the acceptance criteria for the hour time horizon in five of the six cases tested, while SCADA data did not reach the acceptance criteria on either of the two cases tested.

7.3 Recommendations for transition to BAU

Some of the actions recommended in developing the forecasting for a given location as a BAU process are listed below:

- Assessing the quality of historic data available and measures to improve future data collection;
- For locations considered for forecasting, establishing a data I/O process so that data for each location can be easily processed through the toolchain;
- Identifying the time horizons where the accuracy thresholds are consistently achieved (based on a validation exercise such as employed here). In general, longer time horizon forecasts (month ahead and above) can be expected to meet acceptance criteria, given adequate data quality, for primaries, BSPs and load customers that exhibit consistent behaviour;
- Considering the possible drivers for the underlying behaviour and features that might explain its behaviour, based on the DSO's domain knowledge;
- Further developing the toolchain to optimise the model with respect to features and training history, as well as integrating the existing two Jupyter notebooks into a single notebook that:
    a. Better supports automation for use on many different locations;
    b. Allows the user to easily select between using default hyperparameters or perform tuning; and
    c. Allows the user to easily optimise the model for different features and training lengths.

Capita DA believes that the above recommendations can be achieved by a team combining DSO domain knowledge and an engineer with Python skills. The data I/O setup may require data engineering skills depending on the DSO's specific preferences in this domain. With respect to developing the toolchain and testing on different locations, Capita DA believes that specialised data science or forecasting skills are useful if not necessary. Python skills are sufficient to start the process, and the team performing this task would over time build the expertise to optimise the toolchain in line with the DSO's requirements.