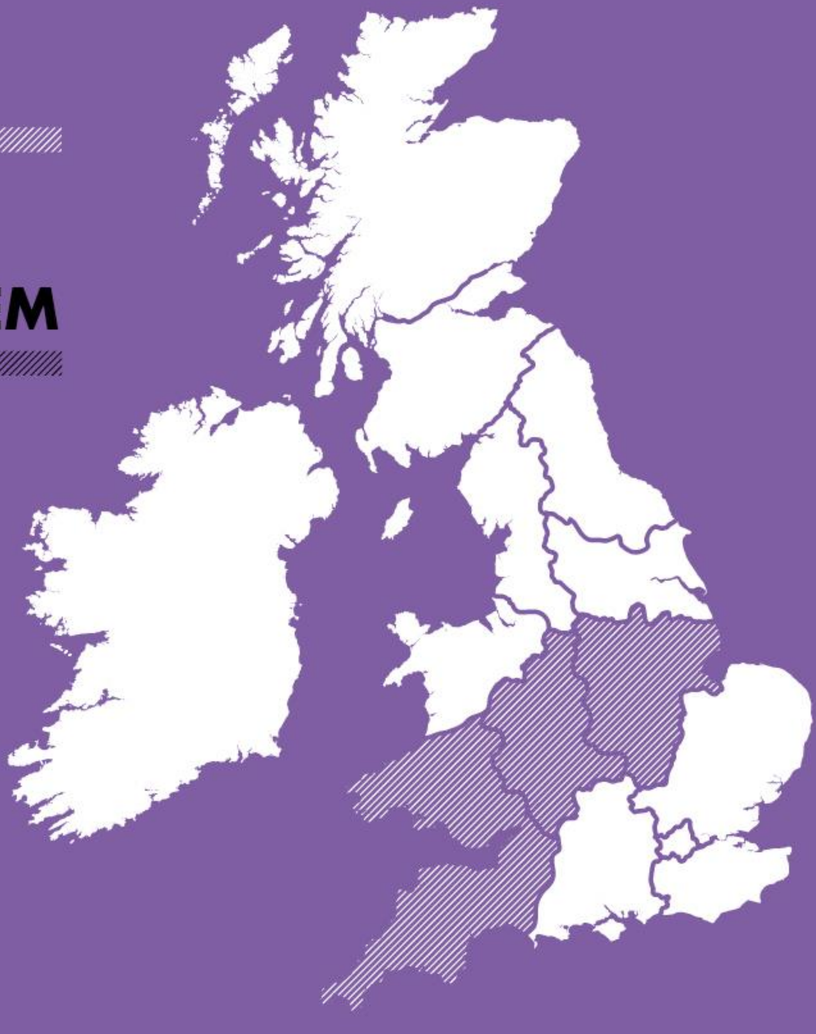
WESTERN POWER DISTRIBUTION

EFFS

ELECTRICITY FLEXIBILITY AND FORECASTING SYSTEM



EFFS
WPD_EN_NIC_003

NIC PROJECT
System Design: Forecasting

RIIO NIC
NETWORK INNOVATION COMPETITION

Report Title	:	System Design: Forecasting
Report Status	:	FINAL REDACTED
Project Reference:	:	WPD/EN/NIC/03
Date	:	25/10/2019

Document Control		
	Name	Date
Prepared by:	Michael Pearson	09/10/2019
Reviewed by:	Jennifer Woodruff	09/10/2019
Approved by:	Ben Godfrey	14/10/2019
	Paul Hartshorne	15/10/2019
	Jonathan Berry	16/10/2019
	Sarah Hadfield	22/10/2019

Revision History		
Date	Issue	Status
18/07/2019	0.1	TEMPLATE
31/07/2019	0.2	DRAFT
05/08/2019	0.3	DRAFT
05/08/2019	0.4	DRAFT
08/08/2019	0.5	DRAFT
16/08/2019	0.6	DRAFT
21/08/2019	0.7	DRAFT
22/08/2019	0.8	DRAFT
17/09/2019	0.9	DRAFT
18/09/2019	0.9.5	DRAFT
03/10/2019	0.9.6	DRAFT
08/10/2019	0.9.7	DRAFT
09/10/2019	0.9.8	DRAFT
09/10/2019	0.9.9	ISSUED FOR APPROVAL
16/10/2019	1.0	FINAL
25/10/2019	2.0	FINAL REDACTED

Contents

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

1 Purpose

The purpose of this design document is to specify how the forecasting requirements defined in the EFFS project's DSO Requirements Specification will be delivered from a functional perspective. This design document forms one of eight system design documents (listed below), namely the forecasting design document. The system design documents complement the System Design Summary Report, which contains an overview each functional area and the relationships between them.

- **Forecasting;**
- Capacity Engine;
- Service Management;
- Optimisation;
- Scheduling;
- Conflict avoidance and synergy identification;
- Market Interface;
- Reporting and Reconciliation.

In accordance with the EFFS Project Direction, this document forms part fulfilment of the project's fourth deliverable to Ofgem, the 'EFFS system design specification'.

2 Executive summary

The project investigated several forecasting methodologies during Workstream 1 (forecasting evaluation and requirements gathering) Smarter Grid Solutions (SGS) was tasked with assessing whether machine learning techniques could perform better than traditional statistical models and Capita Data Science team were tasked with providing independent validation of the proposed models and findings. This work covered a range of assets such as transformers at Grid Supply Points (GSPs), Bulk Supply Points (BSPs), Primary substations and large load customers, but also wind and solar generation sites. Forecasts for time horizons from day-ahead to six-months ahead were considered and the impact of including or excluding certain model features, such as different types of weather data, were tested.

Three forecasting approaches were evaluated. This work concluded that while all the techniques were capable of being tailored to provide reasonable forecasts, the machine learning technique XGBoost gave the best overall balance between accuracy of the results and the effort required to set up and maintain the forecasts. It also found that, as expected, input data quality was an important factor in the quality of the forecast. Surprisingly, while the inclusion of historic weather data improved the quality of the forecasts, they were able to perform reasonably well without this data. This was thought to reflect the model determining seasonal variations from the week-of-year and month-of-year features as a proxy for weather data. However, it was anticipated that day-ahead and week-ahead forecasts would benefit from inclusion of weather forecast data, especially when the predicted weather would be different from the seasonal averages. Similarly, while time-series forecasting methods can be used for forecasting wind or solar generation, they are not the recommended method. While a time-series model may be able to determine a general relationship between the weather data and the generation output, engineering models can better represent the non-linear impacts introduced by inverters, protective control systems etc. and so are the recommended method to model wind and solar generation.

During the system design phase, the focus shifted from which model to implement to the practical questions around implementing the forecasting algorithm. The questions addressed and the conclusions are summarised in the following sections.

2.1.1 Data Sources

The forecasting tool will be provided with historic time-series data from existing SCADA monitoring. This is held within PowerOn[1] and is routinely exported. Existing forecast weather data for solar and wind values will be used of to provide inputs for modelling. The model requires additional forecast data for temperature and for the actual weather experienced. This data, along with a bulk set of historical recorded data for training the model, could be provided by Meteogroup[2] at the same level of disaggregation as the current forecasts.

The design has identified the need for a process step for data correction of the time-series data from SCADA. Further work is planned for the build phase to determine how this can be automated and the degree to which data correction or substitution can be automated. Options for data substitution have been explored and will be examined further in the next phase. Similarly, the work to assess the potential impact of data quality has begun with a review of the monitoring points at a representative location in the trial area.

[1] https://www.gegridsolutions.com/products/brochures/uos/PowerOn_Control.pdf
[2] https://www.meteogroup.com/sites/default/files/180807_weather_data_api_-_corp_factsheet_1.pdf

2.1.2 Accuracy monitoring

The accuracy metrics suggested by the SGS / Capita Data Science team work have been adopted for continuing use within EFFS. There are two purposes to calculate accuracy metrics for the forecasts during the EFFS project. Firstly, it will increase the opportunities for learning what factors affect forecasting accuracy by providing a much larger set of results to analyse. The second reason is to see how forecasting accuracy changes over time.

2.1.3 Forecast locations and aggregation/ disaggregation

While the forecasting algorithm was tested at a variety of locations, the PSS®E[3] power flow analysis software has the capability of aggregating power flows at one voltage level to determine the impact at a different voltage. Therefore, there is no requirement to forecast the load at a BSP transformer if the loads at all the downstream primary transformers and any customers directly connected to the 33kV networks can be forecast. Similarly, PSS®E can manage the aggregation to create a load profile for GSP transformers given the profiles of the relevant BSP transformers and any 132kV connected customers. The same process would apply to 66kV networks. While this reduces the total number of sites that require direct forecasts, it may still be useful to create a small number of forecasts at BSP or GSP transformers for validation purposes.

2.1.4 Forecasting adjustments

The time-series data that is used by the forecasting algorithms is expected to reflect network loadings for standard running arrangements. These are expected to occur for the majority of the time, and non-standard loading values due to maintenance or unplanned outages would be highlighted as outliers by the data cleansing process. However, the most onerous conditions for the network are more likely to be experienced when the network is abnormally arranged and therefore there is a need to adjust the forecasts accordingly. Forecasting the load for these non-standard arrangements using the same forecasting algorithms that are used for normal running is not practical due to the difficulties of identifying when the required running arrangements happened in the past to select the appropriate data, but also because the number of data points would be small, and in discontinuous blocks. Adjusting the load values that include the impact of embedded generation to reflect the total demand when generation is disconnected requires modelling of the embedded generation downstream of a primary substation. As generation can be added to nodes within the PSS®E model, the approach of creating virtual generators, aggregating embedded generation of different types at the primary busbars has been adopted. A similar correction factor is required to reflect that the load on each transformer at a multi-transformer site feeding different busbar sections that are joined by a bus-section circuit breaker will be different according to whether the bus-section is open or closed. A method for estimating the proportional split has been devised based on the expected aggregated load of the outgoing feeders for each busbar section.

2.1.5 Technical implementation

The forecasting evaluation performed by SGS and Capita Data Science team proposes an open source toolchain which could be used to create forecasts. AMT-SYBEX will instantiate the XGBoost forecasting model into their Affinity Networkflow[4] product. To support the trials, Affinity Networkflow will deployed [redacted] to obtain time series and weather data. The output of the forecasting function will provide data inputs to the capacity engine function.

---

[3] https://new.siemens.com/global/en/products/energy/services/transmission-distribution-smart-grid/consulting-and-planning/pss-software/pss-e.html
[4] https://www.amt-sybex.com/networkflow/

3 Glossary

Term	Definition
API	Application Programming Interface
ARIMA	Auto-Regressive Integrated Moving Average is a classic statistical modelling approach for building time-series forecasting models
Constraint	This refers to thermal network constraints (as opposed to voltage constraints)
CHP	Combined Heat and Power
Datalogger	WPD tool for storage of historic time series data, i.e. the data gathered by SCADA systems to support the control room.
DSO	Distribution System Operator
EAC	Estimated Annual Consumption – A value produced by Meter Operators and shared with DNOS representing the expected annual consumption for non half-hourly metered customers
EFFS	Electricity Flexibility and Forecasting System
HISTAN	Historical Analogue file – a record of HH time series data generated from PowerOn currently used to populate Datalogger or TSDS
HH	Half Hourly electricity metering
kV	Kilovolt
kW	Kilowatt
kWh	Kilowatt hour – unit of energy consumption used for billing purposes. Values in Durabill reflect Kilowatt hours which is different to the units used by SCADA (MVA, MW, MVAr)
Long term	Within the context of EFFS, this refers to an activity between six months ahead and one month before the event in question
LSTM	Long Short-Term Memory (LSTM) Artificial Neural Networks is a specific type of deep-learning neural network for learning patterns in time series data
Medium term	Within the context of EFFS, this refers to an activity between one month ahead and the day before the event in question
MAPE	Mean Absolute Percentage Error
NIC	Network Innovation Competition
Affinity Networkflow or	Proprietary software suite developed, licenced and maintained by AMT-SYBEX relating to the management of flexibility services for electricity networks

Term	Definition
Networkflow	
MW	Megawatt
MVAr	Mega volt-ampere reactive
Network hierarchy	The relative configuration of the key locations of the network by voltage level. This is simpler than the integrated network model but would allow an understanding of how actions at a particular primary, for example, would impact on 33kV feeders, bulk supply points, 132kV feeders and GSPs
PowerOn	WPD's Distribution Management System provided by GE
Profile Class	A category for non half-hourly metered customers used to define their consumption profile.
Profile Coefficient	Information on load profiles
PSS®E	Transmission planning and analysis software provided by SIEMENS
SCADA	Supervisory Control And Data Acquisition
SGS	Smarter Grid Solutions (Project forecasting partner)
TSDS	Time Series Data Store
User	Users of the EFFS system will be: • **Forecaster and flexibility co-ordinator** up until the real time management, dispatch and monitoring. Note: both these roles do not currently exist but are required, as they do not map onto an existing business function. The flexibility co-ordinator role will have a very similar skill set to that of an outage planner, whereas the forecaster role will require individuals with a mathematical / statistical background and possibly some programming experience. • **Control engineer** for real time dispatch and monitoring of the network. • **System administrator** system and interface support, maintenance of master data, data cleansing.
WPD	Western Power Distribution
WS1	Workstream 1 of the EFFS project (forecasting evaluation, co-ordination and requirements)
XGBoost	Extreme Gradient Boost a machine learning algorithm used for timeseries forecasting.

4 Related documents

Ref	Document title	Version	Date issued	Prepared by	Location
1	Revised_EFFS_FSP_Redacted_v 2	2.0	06/07/2018	EFFS	Link
2	WPD_EFFS_DSO Requirements Specification_v1.0	1.0	24/05/2019	EFFS	Link
3	WPD EFFS_Forecasting Evaluation Report_v1.0_FINAL	1.0	06/06/2019	Smarter Grid Solutions	Link
4	System Design Summary Report	2.0	25/10/2019	EFFS	Link

5 System overview

5.1 Core functions overview

Figure 1 below is a diagrammatic representation of the functional areas within the EFFS project. The area that is the subject of this document is highlighted in red.

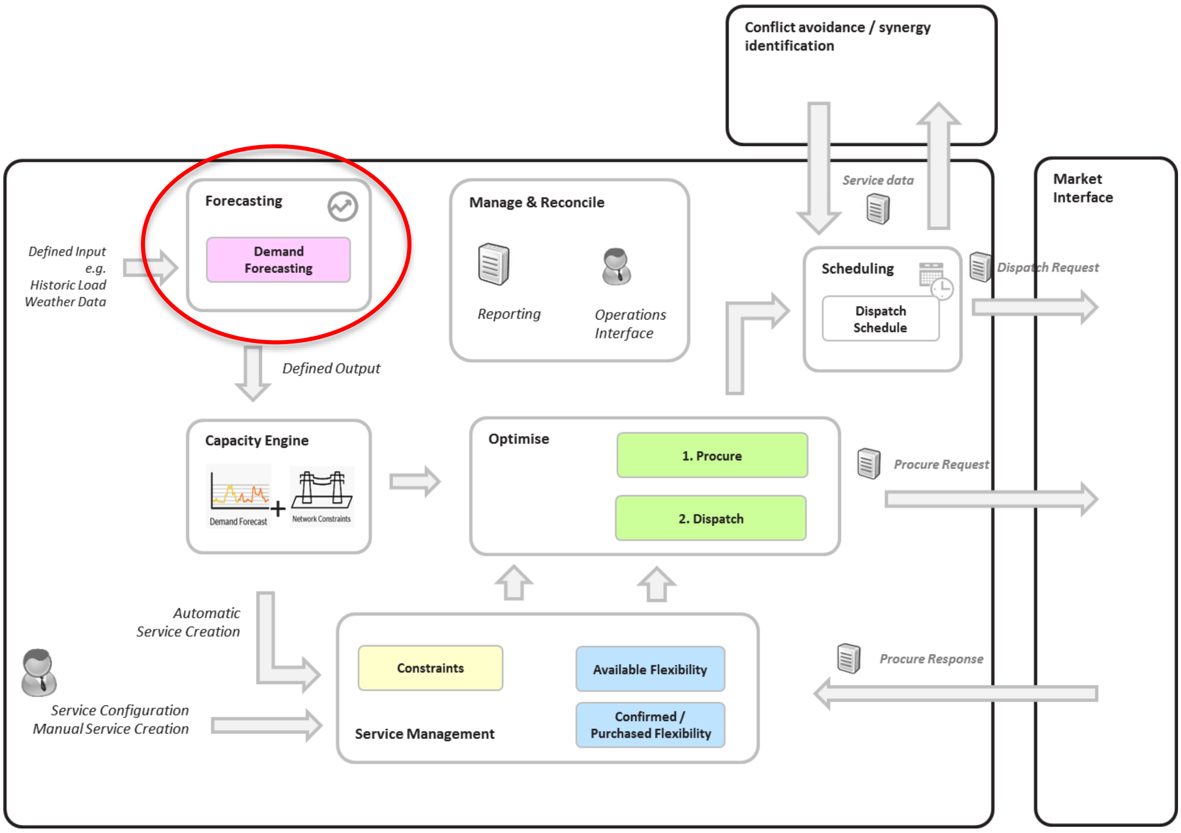


Figure 1: EFFS core functions

6 Forecasting

6.1 Scope

Table 1 defines the scope of the forecasting solution to be implemented in the EFFS project.

In scope	Out of scope	Currently out of scope but to be explored in additional forecasting work
• Implementation of the XGBoost forecasting model provided by Smarter Grid Solutions (SGS) as part of the EFFS project's Workstream 1 (WS1); • Demand forecasting; • Generation forecasting; • Data feeds (historic weather, location, historic load); • Short to medium term forecast horizons (i.e. within day, day ahead, week ahead); • Month ahead to 6-month advance forecast horizons (may not be used but to be explored); • Data handling and data processing as a precursor to use in the forecasting algorithm; and • Conversion of net load to demand.	• Long term forecast horizons (i.e. greater than 6 months); • Long term investment planning; • Implementation of other forecasting solutions or existing models; • Aggregation / disaggregation of forecasts for a: 33kV connected load customer, 33kV generation customer, 132kV connected load customer or 132kV connected generation customer; and • Automated interface for creation of network equipment, though a one-off upload is supported.	• Use of forecast weather data in forecasting* *please see section 9 for further details on these scope items.

Table 1: Scope for forecasting

6.2 Description

The forecasting function in EFFS will see the implementation of a forecasting process that provides demand and generation forecasts for multiple time horizons. Forecasts will be in the form of megawatt (MW) and mega volt-ampere reactive (MVAr) for locations required to support power flow analysis of the 33kV and 132kV networks. This will also be adaptable to include 11kV feeders that would be used to transfer load after the loss of a primary transformer. The forecasting model to be implemented within EFFS has been defined by SGS during WS1 (see 'WPD EFFS_Forecasting Evaluation Report_v1.0_FINAL' for details, summarised below).

The work carried out by SGS during EFFS WS1 explored three different forecasting methodologies:

- Auto-Regressive Integrated Moving Average (ARIMA) – a classic statistical modelling approach for building time-series forecasting models;

- Long Short Term Memory (LSTM) Artificial Neural Networks – a specific type of deep-learning neural network for learning patterns in time-series data; and

- Extreme Gradient Boosting (XGBoost) – a machine-learning technique based on decision trees.

The report also recommended the use of engineering models to forecast the output of wind or solar generation sites. These make use of known performance characteristics of different manufacturer's equipment for given levels of irradiance or wind speed. Open source versions of these models can be acquired (for example, engineering models are available from renewablesninja.com) and utilised within the forecasting module and are expected to be more accurate than the forecasts using time series data.

The detailed findings are captured in 'WPD EFFS_Forecasting Evaluation Report_v1.0_FINAL', but in summary, XGBoost performed the best out of the three forecasting methods in terms of accuracy and usability. Therefore, the XGBoost method has been selected for use in the EFFS trials. The XGBoost model will be instantiated into AMT-SYBEX's Affinity Networkflow product.

Forecasting toolchain
As part of the EFFS project's WS1, SGS developed a forecasting toolchain set; please see 'WPD EFFS_Forecasting Evaluation Report_v1.0_FINAL' for further information. The toolchain provides forecasting methods, which can be used to drive power systems analyses to determine the effect of load and generation on a network i.e. circuit flows through load flow analyses.

For the purpose of this design document, the inputs and outputs of forecasting are defined not the logic and algorithmic processing provided by the toolchain; these are defined in the aforementioned 'WPD EFFS_Forecasting Evaluation Report_v1.0_FINAL'. Throughout this design document, the "toolchain" will be referred to interchangeably between forecasting toolchain/algorithm and Networkflow, owing to its instantiation into AMT-SYBEX's Affinity Networkflow product to support the EFFS trials.

6.3 Solution

This section details how the forecasting function in EFFS has been designed to support the trials.

6.3.1 Pre-requisites

The following pre-requisite is needed for the forecasting function:

- Availability of relevant historic data of sufficient quality to generate reliable forecasts. There are two potential problems here:

  1. Not all sites have monitors recording MW and MVAr values; and
  2. Recorded datasets may not be of sufficient quality which may reflect a variety of different issues.

  To mitigate these, a manual process will be used to create MW and MVAr values from available data for Volts and Amps where these are not recorded directly. A data cleansing process will also be implemented to ensure that any source data used for forecasting are complete and accurate before being passed into the forecasting algorithm. Options for data cleansing and data substitution will be defined during the project's build phase; options for data substitution are presented in section 6.3.4.1.

6.3.2 Input

The forecasting function requires the following inputs:

Input	Description	Primary source
Historic time series data for all values being forecast i.e. load and generation (where this is not expected to be forecast using an engineering model), half-hourly average MW and MVAr.	Approximately 2 years' worth based on the use cases in the SGS work and also to support data cleansing / substitution). This data will be provided for: • Primary substation busbars at 11kV; and • Any customers directly connected to the 25kV, 33kV, 66kV or 132kV networks.	PowerOn
Historic weather data for irradiance, windspeed and, if available, temperature.	These should be provided at no less than hourly resolution. The EFFS forecasting requires HH data so if hourly data is provided then data cleansing / interpolation will be required. The geographic granularity (currently BSP) of the data can be altered without affecting the way in which the algorithm works, but it is expected that having a larger number of more granular weather forecasts would improve the accuracy of forecasting, especially that for renewable generation.	Meteogroup
Forecast weather data (for short term forecasts up to week ahead only).	As for the historic weather data this should be at least hourly resolution and include irradiance, wind speed and, if available, temperature. The EFFS forecasting solution requires Half Hourly (HH) data. If hourly data is provided, then data cleansing / interpolation will be required.	Meteogroup
A lookup table relating each forecast location to the relevant weather monitoring location.	As per the input.	Manual mapping
Capacities of each type of generation to be modelled in aggregate embedded below a primary transformer.	As per the input.	CROWN generation report

Allocation factors for each primary busbar section.	This will reflect the expected proportion of load / generation for the combined HV feeders relating to each busbar covering 48 HH periods for the representative days to be modelled in PSS®E.	As per the below item
Data to support the calculation of allocation factors for each primary busbar.	This may be which may be: 1. Average values for the combined loads as measured by SCADA for all HV feeders associated with a busbar section for representative days, where data quality supports this; and 2. Estimates of the values of combined loads for all HV feeders associated with a busbar section for representative days. These estimates are to be based on the combined Estimated Annual Consumption values and profile coefficients for each profile class for customers associated with each Feeder along with average profiles for half-hourly metered customers.	1.PowerOn 2.Industry average values

Table 2: Inputs for forecasting

6.3.3 Output

The forecasting function will produce the following outputs:

- Forecasts of load and generation as half-hourly average MW and MVAr values for each forecasting location configured in the system (see section 6.3.7);
- Forecasts of the output of embedded generation necessary to convert from net load to demand (see section 6.3.8.16.3.8.1); and
- A set of accuracy metrics (see section 6.3.10) will be produced for each forecast.

The forecasts will conform to the following conventions reflecting WPD policy:

1) In terms of direction of power flow anything outgoing from the busbar is “+” and anything going into the busbar is “-”; and
2) The convention for reactive power is: *“If the current lags the voltage by less than 180o the reactive power is positive. If the current leads the voltage by less than 180o the flow of reactive power is negative”.*

These outputs will be amended in the follow ways:

1. Perform the net load to demand conversion;
2. Select between net load and demand forecasts as appropriate for the contingency being modelled; and

3. Convert forecasts for primary transformers to forecasts for primary busbar (see section 6.3.8.2).

In order to do this the following data is required:

Data	Primary source
Capacities of each type of generation to be modelled in aggregate embedded below a primary transformer.	CROWN generation report
Allocation factors for each primary busbar section reflecting the expected proportion of load / generation for the combined HV feeders relating to each busbar covering 48 Half Hourly (HH) periods for the representative days to be modelled in PSS®E.	As per the below item
Data to support the calculation of allocation factors for each primary busbar, which may be: 1. Average values for the combined loads as measured by SCADA for all HV feeders associated with a busbar section for representative days, where data quality supports this; and 2. Estimates of the values of combined loads for all HV feeders associated with a busbar section for representative days. These estimates are to be based on the combined Estimated Annual Consumption values and profile coefficients for each profile class for customers associated with each Feeder along with average profiles for half-hourly metered customers.	1.PowerOn 2.Industry average values
Lookup of normal busbar running arrangements for each primary.	PSS®E / PowerOn

Table 3: Data required for amending forecasting outputs

6.3.4 Data handling and cleansing

One of the fundamental pre-requisites for the forecasting function in EFFS is that the data is cleansed ready for consumption. The data handling and cleansing process will be managed outside of Networkflow and stored in a database for cleansed data ready to be consumed by the forecasting module. The options for sourcing time series data were:

1. PowerOn holds records of the SCADA values from February 2014;
2. Datalogger, which is the repository of the SCADA data that is stored in PowerOn;
3. TSDS which will fulfil a similar function to Datalogger (and may eventually replace it), this implementation is in progress;
4. The SCADA system itself, though this would duplicate many of the functions already provided to process the data and present it to PowerOn; and
5. The HISTAN (HISTorical ANalogue) files which are produced to transfer POF data between PowerOn and Datalogger. For reference an example HISTAN file is included in Appendix 2.

Initial expectations were that Datalogger would be the best source for this data as it could be interrogated with no potential impact to PowerOn. Moreover, it was presumed that Datalogger may have a longer time series than was held within PowerOn and that having more previous years of data would be needed to support forecasting. However, it has become clear that:

1. PowerOn stores significant volumes of data representing multiple years;
2. Time series of greater duration than 2 years does not necessarily improve forecast quality;
3. Avoiding the translation from PowerOn to Datalogger may make it easier to relate time series data to the relevant assets within EFFS. Matching data loggers back to assets is a difficult process and having one fewer lookup table to navigate may be beneficial; and
4. The process to create the export file, normally compiled overnight to export data from PowerOn to Datalogger may provide a useful means to export data from PowerOn more frequently without requiring modifications to PowerOn by General Electric.

Therefore, the decision was made that the initial provision of historical time series data will be a one-off exercise querying the historic data held in PowerOn. The ongoing routine updates of time series data will be taken from the HISTAN files generated between PowerOn and Datalogger, as this will leverage the same process of matching analogues to assets as in the initial upload bulk upload.

Currently WPD is working towards full implementation of TSDS by the end of 2019. This should not cause any issues for EFFS as both systems rely on the same file format (HISTAN) that are created on a nightly basis. A process runs on each DNO system that creates these HISTAN files and copies them to a directory called \datafiles\histan.

There is currently a separate process investigating if there is a need for the data to be cleansed so there may be an intermediary database that the HISTAN files will be loaded into so data cleansing can take place first then the data will be loaded into Networkflow. As the HISTAN files are currently created on a nightly basis and uploaded to Datalogger there will be minimal impact on the current process to take these HISTAN files and upload them to the intermediary database for cleansing. Therefore, the proposed process for providing on-going time series data for use by the forecasting function and evaluation of the forecasting accuracy is as follows. An overnight process on a configurable timer is run which causes the HISTAN file to be imported and processed by the EFFS data cleansing database and cleansed data to be produced in tables that are used by the EFFS forecasting tool for both forecast production and evaluation. There will be no extraction directly from Datalogger or TSDS.

Potential issues
It is known that the matching the analogues to the assets within PSS®E may be an issue due to the lack of a common identifier. We need to match PSS®E assets to PowerOn assets for two different purposes within EFFS:

1. for matching forecasts based on PowerOn data to the relevant assets in the PSS®E model; and

2. to match the position of switches in the PowerOn model to those within the PSS®E model.

Therefore, being able to match these assets together is beneficial for both uses.

6.3.4.1 Data substitution options

Below are data substitution options that the project will progress.

Primary Busbar

Step 1: When creating forecasts for each Primary Transformer the following data sources will be used in order in order of preference:

- First preference: use loggers for MW and MVAr;

- Second preference: use loggers for kV and Current and assume power factor; and
- Third preference: for aggregate of customers connected to HV feeders for each busbar section – create estimate based on half hourly average meter readings and EAC x profile co-efficient for each profile class, convert from kWh to power with assumed power factor.

<u>25kV, 33kV, 66kV or 132kV connected Wind or PV site</u>

Below are the possible approaches for how to forecast for a 33kV, 66kV or 132kV connected Wind or PV site. It is still to be determined which option is most preferable.

- First option – use engineering model and weather forecast to create power forecast – assume power factor to obtain MW and MVAr;
- Second option – use MW and MVAr loggers to create time series forecast with XGBoost;
- Third option – use kV and current loggers to calculate power, assume power factor to create MW and MVAr values, use these to create time series forecast with XGBoost; or
- Fourth option – relate site to MPAN for HH metering data, get MWh and MVArh export data from Durabill, convert to MW and MVAr, use time series forecast algorithm with XGBoost.

<u>25kV, 33kV, 66kV or 132kV connected other generation or load site</u>

Same as above but no option for engineering model.

6.3.5 Weather data

EFFS will use several sources for weather data. These will include two types of group, historic and forecast data.

Historic data
Historic data will be received from the following source:

- Two years historic and ongoing actuals from Meteogroup.

It will contain the following data items:
- 'Irradiance';
- 'Temperature'; and
- 'Wind speed'.

Forecast data
Forecast data will be received from the following sources:
- Seven day for wind and solar from internal WPD sources; and
- Seven day forecast for temperature from Meteogroup.

It will contain the below data items:
- 'Irradiance';
- 'Temperature';
- 'Wind speed'.

Weather data mapping to network
The forecasting function in EFFS will require a lookup table to determine how the historic and forecast weather locations map to the network locations. This is currently achieved based on BSP grouping. The table will be maintained and populated manually by a user and then by the solution to

determine the mapping of the weather data to the relevant network location based on the data in the table.

6.3.6 Create equipment

In Networkflow, 'Equipment' can be created for several nodes such as a BSP, primary substation or other entity used in the network hierarchy. This enables the system to forecast and procure flexibility requirements at the appropriate level of network hierarchy. Therefore, it will be necessary to create an equipment record for every node involved in these processes.

Equipment will be created using the standard Networkflow "Create Equipment" interface. This is an API and may be called automatically from an external system or manually through a tool such as a UI. Note: for the purposes of the EFFS trial, equipment may also just be created manually via direct DB access. A one-off activity will be conducted to extract and transform the existing PSS®E network hierarchy into the inserts required for a SQL Script. Once an initial hierarchy has been inserted into Networkflow, a system admin user will maintain the data integrity of the hierarchy such as when a change of network configuration has occurred.

The creation of equipment will include the creation of parent child relationships between pieces of network equipment. Therefore, were a system admin user wishes to reflect the hierarchy such as the relationship between the 132kV (parent) and 33kV (child) network the user can include the parent equipment ID to each child equipment ID. This will be fully managed by a system admin user as no network import functionality exists.

Details on how this is to be done in Networkflow are captured in Appendix 1.

PSS®E/PowerOn lookup table
EFFS deems the PSS®E ID to be the master Equipment ID for Networkflow. A manual one-off exercise will be required to create a lookup table to match the PowerOn to PSS®E ID where they do not match. This will build on the experience gained during project Equilibrium where a similar exercise was required for the voltage optimisation system. The table will then be maintained manually by a system admin user. The administrator will perform a routine check for upcoming changes to the network on a fortnightly basis by checking for additions to the forward-looking "committed" network model. These additions are likely to be added to the "committed" model before network patches are created for use within PowerOn. This will prompt the administrator to find the relevant patches in PowerOn and their IDs can then be included in the lookup table ready for use when the patches go live. The relevant equipment can then be created in advance in Networkflow to provide an equipment ID and vice versa for any other system that has misalignment of IDs between systems. The lookup table will be held in the data cleanse database and maintained by a system administrator and other systems will interface if required to determine which ID to use for the respective system.

6.3.7 Forecasting types

The forecasts to be produced for each type of equipment / site type are detailed in Table 4 below.

Site type	Forecast quantities
25kV connected load customer	MW, MVAr
33kV connected load customer	MW, MVAr
33kV connected generation customer	MW, MVAr

66kV connected load customer	MW, MVAr
66kV connected generation customer	MW, MVAr
132kV connected load customer	MW, MVAr
132kV connected generation customer	MW, MVAr
Primary transformer	MW, MVAr
Busbar section	MW, MVAr

Table 4: Forecast types

In terms of direction of power flow as per current WPD policy, anything outgoing from the busbar is "+" and anything going into the busbar is "-".

6.3.8 Forecasting Adjustments

6.3.8.1 Net load to demand conversion

The load drawn at primary transformers and used for time-series forecasting, will reflect the impact of generation that is embedded in the 11kV and LV networks. The monitoring records the net load which is the demand of the network less that which has been satisfied by the output of embedded generation. However, for modelling the network after a fault, when generation within the fault zone would be disconnected from the network by protective relays, then the most appropriate value to use for modelling within the fault zone would be the demand, rather than the net load. This demand has to be estimated as it cannot be measured directly. A number of potential correction mechanisms have been considered, including:

- Model the Distributed Generation analogue which is populated in Datalogger and add this value to the net load forecasts;
- Model each instance of distributed generation downstream of the primary busbar and add the combined value to the net load forecasts; and
- Model downstream distributed generation as virtual aggregated generators for each type of generation as if they were located at the primary substation.

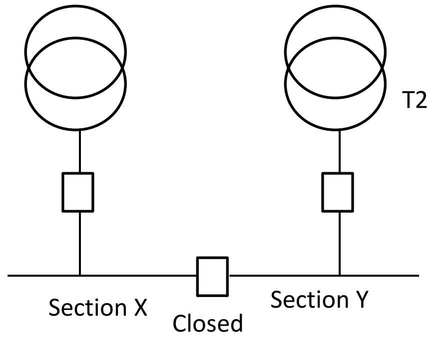
The selected option is to create virtual aggregated generators for each type associated with each primary busbar section and to forecast the generation that they would output. This should reflect the impact of different weather conditions more clearly than using the combined analogue in Datalogger. Modelling each installation individually is unlikely to improve accuracy unless more spatially granular weather data is provided.

For wind and solar generation, the default engineering models (to be determined by the additional forecasting work) will be used alongside the weather forecast data. The combined capacity will be given by the known capacity of each generation type associated with each 11kV busbar section. There is a risk that the known capacity underestimates the actual capacity as many domestic installations are not correctly registered with DNOs. These forecasts will be compared to the value in Datalogger for distributed generation for the whole primary to determine whether a scaling factor needs to be applied. For other forms of generation known to be connected, e.g. Combined Heat and Power (CHP) or anaerobic digestion, an average profile will be determined from data available in Durabill which will then be scaled by the connected capacity.

6.3.8.2 Mapping primary transformers to busbars for modelling

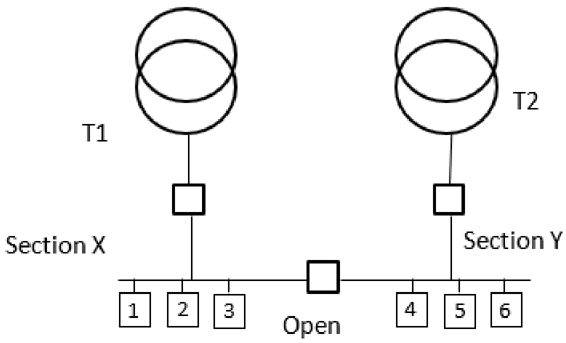
EFFS will forecast load at transformers using the time-series data relating to those transformers. However, the PSS®E model requires load/generation values to be associated with the 11kV busbars, rather than transformers. The following process is used to translate between transformer and busbar loadings.

1. Where a busbar / busbar section is normally run with only one transformer connected then the busbar load/ generation is equal to the transformer load/ generation.
2. Where multiple transformers are normally connected to a busbar / busbar section then for modelling the busbar under *normal running* the load/generation is equal to the combined load/generation.



Bus section between sec. X and sec. Y is normally closed. Busbar sections x & y are treated as a single busbar within PSS®E with busbar load = Load for T1 + Load for T2

3. Where multiple transformers are normally connected to a busbar / busbar section then for modelling the busbar under conditions where the bus section *is now open,* e.g. planned outage, then the combined load of the connected transformers needs to be split. The option of creating an adjustment by using monitoring data from the bus-section was discounted due to the issues around the coverage of this data and the certainty of the direction of current. The combined load should be apportioned according to the sum of the expected loads on the 11kV feeders relating to each busbar section. In the example below the load at section X = Load for T1 + Load for T2 x (sum of load on feeders 1-3) / (sum of load on feeders 1-6).



Estimates for the loads on each feeder can be generated from combined Estimated Annual Consumptions for each profile class and average profiles for half hourly metered customers. Proportions of load for each section of busbar will need to be estimated for each half-hour but will be averaged for each season rather than needing to be generated for every day of the year.

6.3.9 Forecasting timelines

The forecasting timelines and operational use cases to be supported are detailed in the table below.

Forecast Horizon	Time range for forecast	Regeneration frequency	Business function
Hour ahead	0 to +1h	Ad hoc	It has been deemed that due to the short-range nature and timing of this forecast it would make better business sense to use the day ahead and the respective contingencies to determine the best course of action in the event a fault occurs. Therefore, it will remain part of the forecasting functionality but not used in EFFS. See section 5 of 'WPD_EFFS_DSO Requirements Specification_v1.0' for details.
Day ahead	0 to +1d	Overnight	Dispatch of scheduled constraint management. See section 5 of 'WPD_EFFS_DSO Requirements Specification_v1.0' for details.
Week ahead	0 to +7d	Overnight	Procure scheduled constraint management, pre-fault constraint management and post-fault constraint management services. See section 5 of 'WPD_EFFS_DSO Requirements Specification_v1.0' for details.
Month ahead	0 to +1 calendar month	Weekly	Validate flexibility already procured is sufficient and forecasting accuracy test to be used in the EFFS trials. This will not be used to trigger any market interaction as none of the flexibility platforms EFFS will interface with operate in these timescales.
Six months ahead	0 to +6 calendar months	Monthly	Validate flexibility already procured is sufficient and forecasting accuracy test to be used in the EFFS trials. This will not be used to trigger any market interaction as none of the flexibility platforms EFFS will interface with operate in these timescales.

Table 5: Forecasting timelines

6.3.10 Accuracy metrics

Networkflow will store each day's forecast values until actual data values are available. The forecast values will be provided via an historic time series data interface. Once received, Networkflow will calculate the accuracy metrics for each Half Hourly (HH) period of each forecast. Accuracy metrics will be calculated as MAPE (Mean Absolute Percentage Error), expressed by the following formula:

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

Accuracy metrics will be available as follows:

Forecast version	Accuracy metrics available for
Hour ahead	Previous 6 months of daily values
Day ahead	Previous 6 months of daily values
Week ahead	Previous 6 months of daily values
Month ahead	Previous year of weekly values
Six months ahead	Previous 2 years of monthly values

Table 6: Accuracy metrics

Figure 3 contains a comparison for day ahead forecasts between forecast and actual HH values taken from the SGS work.

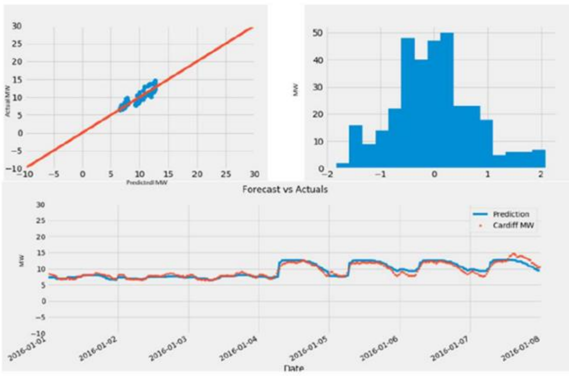


Figure 2: Forecasting accuracy monitoring

6.4 Changes since DSO requirements document baselined

Based on the learnings from the SGS forecasting evaluation report and the WPD experience of power flow analysis in PSS®E, it has been determined that forecasts at lower voltage levels are most appropriate to feed into the analysis carried out within the capacity engine. These inputs are then aggregated to higher voltage levels within the PSS®E package. Therefore, GSP and BSP forecasts are not required. Moreover, having forecasts at this level of granularity is especially beneficial because there is no need to alter the forecasts when the network configuration is non-standard, but rather the amended power flow can be calculated in PSS®E easily.

7 Interfaces

Figure 2 provides an overview of the interfaces to be implemented in support of the EFFS forecasting solution.

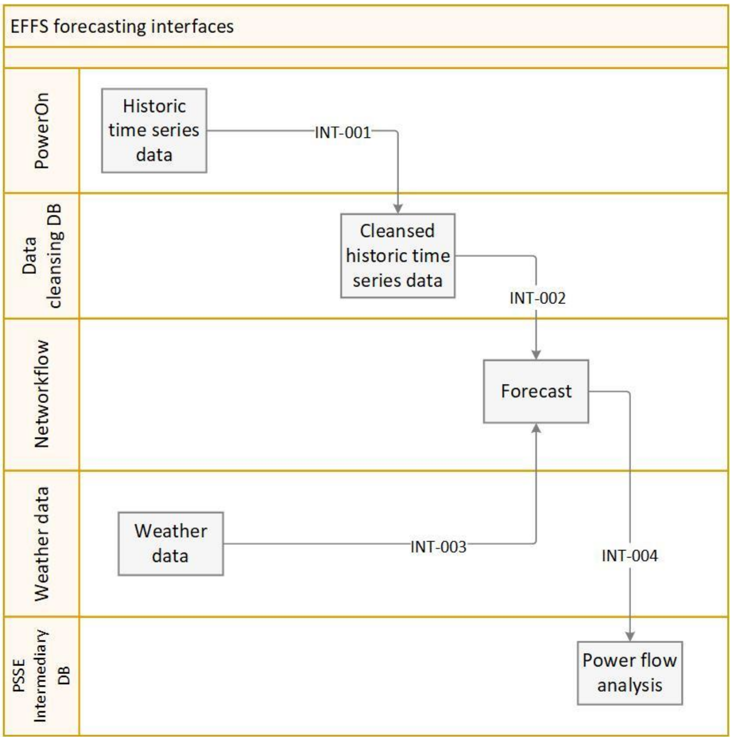


Figure 3: EFFS forecasting interfaces overview

Details on the four interfaces are specified in Table 7 below.

Interface	Source system	Target system	Type	Frequency	Data
INT-001	PowerOn	Data Cleanse DB	[redacted]	Daily	Raw historic HH load and generation data for all equipment where a forecast is required.
INT-002	Data Cleanse DB	Networkflow	[redacted]	One off initial migration tranche, daily updated of uncleansed data and monthly refresh of cleansed data.	Cleansed historic HH load and generation data for all equipment where a forecast is required.

Interface	Source system	Target system	Type	Frequency	Data
INT-003	Weather data	Networkflow	[redacted]	Daily	Historic and forecasted weather data
INT-004	Networkflow	PSS®E Intermediary DB	[redacted]	Daily	Forecast load and generation data for selected equipment.

Table 7: Forecasting interfaces

8 Data items

The following section lists the data items to be contained in the interfaces described above.

The interfaces are described in an indicative logical fashion rather than physically as this information is proprietary. The detailed physical interfaces will be agreed during the build phase of EFFS.

8.1 INT-001 Raw historic time series data

Data item	Type	Units	Cardinality	Valid set value	Notes
Transaction type	VARCHAR(50)	N/A	1	'Historic time series data'	
Transaction ID	NUMBER(10)	Numeric	1		Unique ID for the transaction. Should be included in any related responses generated by the generating system.
Transaction Datetime	TIMESTAMP	TIMESTAMP	1		Date and Time when the request was created in the following format 'YYYY-MM-DD HH24:MI:SS.FF'
Equipment ID	VARCHAR(14)	N/A	1		Unique identifier for the equipment. It is expected that the PSS®E ID will be used and then a lookup table will be used to match PowerOn to PSS®E IDs.
Unit of measurement	VARCHAR(4)	N/A	1	'MW' 'MVAr'	
HH Datetime	TIMESTAMP	TIMESTAMP	1-*		This will be defined by a DATE + TIME of the end of the HH period.
Value	NUMBER(10,3)	Numeric	1-*		

Table 8: INT-001 data items

8.2 INT-002 Cleansed historic time series data

As per INT-001.

8.3 INT-003 Weather data

Data item	Type	Units	Cardinality	Valid set value	Notes
Transaction type	VARCHAR(50)	N/A	1	'Historic time series data'	
Transaction ID	NUMBER(10)	Numeric	1		Unique ID for the transaction. Should be included in any related responses generated by the generating system.
Transaction Datetime	TIMESTAMP	TIMESTAMP	1		Date and Time when the request was created in the following format 'YYYY-MM-DD HH24:MI:SS.FF'
BSP Group ID	VARCHAR(14)	N/A	1		Unique identifier for the BSP Group. This will be linked to the Equip ID via a lookup table between the two held in Networkflow.
Weather Type	VARCHAR(10)	N/A	1	'Historic' 'Forecasted'	
Measurement type	VARCHAR(15)	N/A	1	'Irradiance' 'Temperature' 'Wind speed'	
Unit of measurement	VARCHAR(15)	N/A	1	'W/m2' 'Degrees celsius' 'km per hour'	
HH Datetime	TIMESTAMP	TIMESTAMP	1-*		This will be

Data item	Type	Units	Cardinality	Valid set value	Notes
					defined by a DATE + TIME of the end of the HH period.
Value	NUMBER(10,3)	Numeric	1-*		

Table 9: INT-003 data items

8.4 INT-004 Forecast output

Data item	Type	Units	Cardinality	Valid set value	Notes
Transaction type	VARCHAR(50)	N/A	1	'Forecast output'	
Transaction ID	NUMBER(10)	Numeric	1		Unique ID for the transaction. Should be included in any related responses generated by the generating system.
Transaction Datetime	TIMESTAMP	TIMESTAMP	1		Date and Time when the request was created in the following format 'YYYY-MM-DD HH24:MI:SS.FF'
Equipment ID	VARCHAR(14)	N/A	1		Unique identifier for the equipment
Forecast parameter	VARCHAR(4)	N/A	1	'MW' 'MVAr'	
Forecast type	VARCHAR(2)	N/A	1	'SM' 'MA' 'WA' 'DA' 'HA'	'SM' = Six months ahead' 'MA' = Month ahead' 'WA' = Week ahead' 'DA' = Day ahead' 'HA' = Hour ahead'
Load / generation	VARCHAR(1)	N/A	1	'L' 'G'	'L=' 'Load' 'G' = Generation'

Data item	Type	Units	Cardinality	Valid set value	Notes
HH Datetime	TIMESTAMP	TIMESTAMP	1-*		This will be defined by a DATE + TIME of the end of the HH period.
Forecast value	NUMBER(10,3)	Numeric	1-*		Will contain a '+' to indicate a positive value, a '-' to indicate a negative value.

Table 10: INT-004 data items

9 Additional forecasting work

The project plans to explore further refinement of the existing forecasting toolchain set to enhance its forecasting ability.

The additional forecasting work aims to achieve the following objectives:

- Explore and test recommendations that have been proposed in the EFFS reports, such as net load to demand conversion and applying engineering models to wind and solar (PV) generation;
- Provide a comparison of Datalogger and Durabill data sources to use for 33kV or 132kV connected customers with respect to identifying the preferred source;
- Simulate load and generation forecasts (using the same time horizons as in the EFFS project) on a sample of locations in the Trial Area and report on findings with respect to key drivers behind forecast accuracy; and
- Evaluate the performance of the forecasting tool for forecasting the load /generation of 11kV interconnector feeders on behalf of the TRANSITION and FUSION projects.

Table 11 outlines the scope of work to be undertaken by the additional forecasting work:

Item No.	Description	Scope	Comment
1	**Trial Area Definition**	From the recommended GSP groupings that have been suggested as needing to be modelled together, generate a list of forecasting points required i.e. primary busbar sections, 33kV connected customers and 132kV connected customers.	To be carried out in-house by the EFFS project team.
2	**Data Quality / Data Cleansing**	Determine processes to assess and correct data quality for the SCADA time-series data.	Some initial work to already carried out by the EFFS team. This will be assessed to determine how this can be applied and automated to the data cleanse database.
3	**Weather data**	Source weather data items and amend the XGBoost model to use forecast data as well as historical data.	Weather data sourcing on-going with Meteogroup. Assistance required to amend the XGBoost.
4	**Wind & PV forecasting using engineering models**	Investigate the engineering models for wind and PV generation as an alternative to XGBoost: ▪ research engineering models for wind and PV; ▪ perform forecasts using the selected models; and ▪ compare accuracy against XGBoost models.	Requires expert assistance. To be outsourced.
5	**Large customers excluding wind or PV generation**	Where these customers have both SCADA and Durabill data: ▪ perform XGBoost forecasts using Durabill data and Datalogger data; ▪ compare the relative accuracies	Requires expert assistance. To be outsourced.

		and establish if one is a clearly preferred source; and ▪ perform forecasts using simple averaging models. Establish the relative performance of simple averaging models vs. XGBoost model.	
6	**11kV feeder modelling**	A selection of 11kV interconnectors will be modelled using the XGBoost algorithm to determine the levels of accuracy.	Requires expert assistance. To be outsourced.
7	**What drives the accuracy / inaccuracy at different sites?**	Using the additional forecasts and accuracy results for items 2-7 above, can further conclusions be drawn on the drivers of inaccuracy in forecasting e.g. the relative importance of data quality, the volume of embedded generation, etc.?	Included in the work to be outsourced.

Table 11: Additional forecasting work

An initial approach to outsource this work has highlighted the need to pre-select sites for the analysis and present cleansed data in order to reduce the uncertainty and perceived risk to the company performing the analysis. Providing a scope that includes an increased level of detail should reduce the risk of misinterpretation but also the costs.

9.1 Forecasting of 33kV and 132kV connected solar and wind generation.

The forecasting work by SGS found that seasonal variations in output for generation could be captured by including features in the model for "month of the year" or "week of the year". Improved accuracy was achieved when historical weather data was included for irradiance or windspeed. This type of model may be suitable for forecasts that are beyond a week in advance, where weather forecasting accuracy can be expected to be low.

However, where weather forecasts may be expected to be useful indicators, such as for next day or week in advance forecasts, then the recommended method from the SGS work for forecasting generation is to use an engineering model which estimates the output for a given level of wind or irradiance.

There are many free-to-use models available which reflect the operating characteristics of different makes and models of generator and can reflect features such as the height of the wind turbine axle above ground level. Given the higher level of expected accuracy of these models, these should be used. Further work is planned to determine what data is available within WPD to select the best model or which model can be used as a default where no information about make/ model can be obtained.

This will be explored in the additional forecasting work but based on the above rationale Table 12 shows the expected forecasting methods for 33kV and 132kV connected solar and wind generation.

Forecast horizon	Forecast method
Day ahead, week ahead	Engineering model with weather forecast data
Month ahead or longer	Time series forecast including historical weather data or engineering model with seasonal average forecast weather data.

Table 12: Forecasting of 33kV and 132kV connected solar and wind generation

10 Contact

If you have any questions relating to this document, please use the following points of contact:

Future Networks Team:

Western Power Distribution,
Pegasus Business Park,
Herald Way,
Castle Donington,
Derbyshire
DE74 2TU

Email: jwoodruff@westernpower.co.uk

Appendix 1 – Networkflow create equipment interface

Table 13 below lists the data items required to create equipment in Networkflow.

This interface is described in an indicative, logical fashion rather than physically as this information is proprietary. The detailed physical interfaces will be agreed during the build phase of EFFS.

Data item	Type	Units	Cardinality	Valid set value	Notes
Transaction type	VARCHAR(50)	N/A	1	'CreateEquipment'	
Transaction ID	NUMBER(10)	N/A	1		Unique ID for the transaction. Should be included in any related responses generated by the generating system.
Transaction Datetime	TIMESTAMP	N/A	1		Date and Time when the request was created in the following format 'YYYY-MM-DD HH24:MI:SS.FF'
Equipment ID	VARCHAR(14)	N/A	1		Unique identifier for the equipment. It is expected that the PSS®E ID will be used and then a lookup table will be used to match PowerOn to PSS®E ID's.
Equipment Type	VARCHAR(50)	N/A	1		
Equipment Name	VARCHAR(50)	N/A	0-1		
Asset Rating	NUMBER(5,2)	Amps	1*		
Location ID	VARCHAR(50)	N/A	0-1		
Geo Location	VARCHAR(20)	N/A	0-1		

Data item	Type	Units	Cardinality	Valid set value	Notes
Parent Equipment ID	VARCHAR(14)	N/A	0-1		

Table 13: Data items to create equipment in Networkflow

Appendix 2 – HISTAN file

Embedded is an example HISTAN file format.

[redacted]

Below is an example HISTAN file.

[redacted]