WESTERN POWER DISTRIBUTION

INNOVATION

NEXT GENERATION NETWORKS

TIME SERIES DATA QUALITY
CLOSE DOWN REPORT

Report Title	:	TIME SERIES DATA QUALITY CLOSE DOWN REPORT
Report Status	:	Final
Project Ref	:	WPD_NIA_011
Date	:	01/03/2017

Document Control

	Name	Date
Prepared by:	James Bennett	20/02/2017
	Paul Charleton	
Reviewed by:	Roger Hey	01/03/2017
Approved (WPD):	Roger Hey	01/03/2017

Revision History

Date	Issue	Status
31/08/2016	v1.00	Final
25/10/2016	v1.1	Final Formatted
20/02/2017	v2.0	Minor Edit
01/03/2017	V3.0	Definitive Issue

Contents

Executive Summary **5**
1 Project Background **6**
1.1 Problem 6
1.2 Method 6
1.2.1 Approach 6
2 Scope & Objectives **7**
3 Success Criteria **8**
4 Details of the Work Carried Out **9**
4.1 Cross Reference Analysis – PowerOn NMC System to Datalogger Offline Data Access System 9
4.2 Data Analysis and Visualisation 10
4.3 Extended Data Analysis 13
4.4 Stakeholder Engagement 18
4.5 Pilot Rectification Programme 18
5 The Outcomes of the Project **18**
6 Performance against Project Aims, Objectives and Success Criteria **20**
7 Required modifications to the planned approach during the course of the project **20**
8 Significant variance in expected costs and benefits **21**
9 Lessons Learnt for Future Projects **21**
10 Planned Implementation **23**
11 Facilitate Replication **23**
11.1 Knowledge Required 23
11.2 Products/Services Required 24
12 Contact **24**
Reference Documents **24**
Appendix 1 – Data Analysis Results **26**

Glossary

Abbreviation	Term
CSV	Comma Separated Values. A file format where data elements are separated by commas on each row and structured in multiple rows for each new record. CSV data is easily imported into inspection and analysis programs such as MS Excel.
CE	Calculation Engine (PowerON element)
CT	Current Transformer
DNO	Distribution Network Operator
Heat Map	A data presentation/inspection method wherein structured/ tabulated data is represented by a colour dependent on the value. The overall range of the data (from minimum to maximum) is divided into a number of subranges each associated with a warmth colour in order by value. The resulting heat map gives a quick and ready way to visualise where the overall maxima and minima and other data features are located.
HH	Half Hour (or Half Hourly) data points, being averages over this interval
HISTAN	Analogue History Database in PowerON
INM	Integrated Network Model, a next generation database tool derived from existing diverse sources of data
PowerON	Power on (Fusion), the WPD NMS system. Also referred to as ENMAC for historic reasons and also sometimes as PoF.
RTU	Remote Terminal Unit
SCADA	Supervisory Control And Data Acquisition. Systems for remote monitoring and control operations via communications channel via which analogue values are returned from the remote sites to the centre (PowerON).
SPSS	IBM Statistics Professional analytics program
VBA	Visual Basic for Applications, Embedded Excel programming language
VT	Voltage Transformer
WPD	Western Power Distribution

Executive Summary

The UK Energy industry is changing and changing very quickly. Technology is not only changing how consumers use electricity and when, but it also gives DNO's access to more data. The question is though, what to do with it to extract the maximum benefit and what is the best way to get the data.

With improved instrumentation and telecommunications infrastructure, the volume, variety and rate of acquisition of data available from the electricity distribution network continues to increase. In order to utilise this data effectively we must quickly and effectively interpret it to understand the relevant information, identify what is incorrect or missing, and extract the underlying value.

In addition to using data for real-time control purposes, WPD keeps historic time series data from a wide variety of sources within a number of databases which can be interrogated on an 'as needed' basis. Generally this would be for planning purposes. Due to the sheer volume of data, errors, omissions and underlying trends are difficult to spot and often rely on manual intervention alone.

The Time Series Data Quality project aimed to improve the quality and usability of the network operational data by developing 'Big Data' analysis techniques, and using these to look specifically at that data collected on a half hourly basis using an initial pilot evaluation area covering the WPD South West region. The main objective being to understand the issues and determine how best to address them. It was also intended that by understanding the data more comprehensively, that this would also benefit our control systems. At the end of the project we would then disseminate our findings for all DNO's to then apply any learning for their own purposes.

One of the first major findings was that traditional data analytics in its 'black box' form cannot be used on the current data set due to the perceived large number of possible, but unqualified inaccuracies contained within the data. This is not an atypical result, many data landscape validation exercises reveal a similar initial conclusion and usually result in a data cleansing programme (which in our case is now underway on the analogue set).

The project team have worked with WPD Control Managers and other Senior Stakeholders to derive a strategy for rectification of a pilot set of data corrections. This would then enable us to scope and better understand the issues raised. We also compiled at the start a master set of analogue related issues and the definition and use of KPIs to track clearance of has activities that would be undertaken.

This Closedown Report presents the final conclusions of the Project. The Closedown Report also provides detail of the tools developed to analyse the 35,000 or so, individual analogues

available from the WPD operating regions and the resulting rectification programme outcomes which are now returning feedback on progress.

1 Project Background

1.1 Problem

WPD stores historic time series data across the network. These are contained within a number of isolated databases which can be interrogated on an 'as needed' basis. This is done generally for planning purposes.

The size of these databases has increased monotonically (see below) as new analogues have continued to be added, while obsolete entries have not been systematically reviewed and removed where appropriate, and cross-database consistency has not been assured. The project sought to investigate the use of data analytics to understand data quality issues so as to be able to identify trends and issues which would not be detected otherwise in normal modes of operation.

1.2 Method

Established 'Big Data' analytics techniques were used to process some of the huge amounts of time series data held, initially for one licence area. This was done in order to identify data quality issues and emerging trends. IBM were engaged to use their SPSS analytics tool to further the investigations which included "number crunching" using readily available desktop analysis tools. At the same time in-house analysis was also carried out using regular business tools. The results of the IBM and in-house operations were compared to validate both workstreams and activities modified to avoid unnecessary duplication.

The developed tools were as generic as possible and the processes and outcomes documented in order to replicate the analysis across the other WPD licence areas and to other DNO's.

1.2.1 Approach

The approach to the work was as follows:

- Create an overview of the data process and identify where the possible sources of error enter the system – obtain an "already seen" list of error types and match to this;
- Determine the signatures of the different types of errors and establish how to identify and flag these (use of tools to do rapid bulk data plots to allow visual inspection, cross reference between the main data management systems/tools, current sum checking, heat map analysis);
- Establish business linkage processes for dealing with e.g. PowerOn teams for obtaining data and determine how to feed-back on and then fix the detected errors/issues;

- Manage the error rectification process and check fixed status using KPIs;
- Linkage to end user of the datasets to verify that the programme of work is successful.

Initially the project focussed on time series data held within the WPD 'in-house' data historian known as 'Datalogger' which sits within the companies asset management software. This time series data consists primarily of half hourly average Amps, Volts, MW, MVAR, MVA data together with smaller datasets on Tap Position, Weather etc. Datalogger receives data through an overnight batch process from the Network Management System, due the potential for errors in this process the data analysis was subsequently expanded to include a data extract from our raw data source 'Histan' held within the Network Management System.

2 Scope & Objectives

Two years' worth of historic time series data for the South-West License area was initially taken from the available databases and used for data analysis. A set of repeatable and scalable processes were established to;

- Gain a better understanding of the data;
- Identify gaps;
- Identify suspect/defective data;
- Explore the creation of rules to replace missing/defective data;
- First detect then assign directions to power flows where absent;
- Identify the causes of suspect data through common patterns.

The project also looked at what other information could be incorporated into 'Big Data' analytics to further validate data quality.

The project high level objectives were as follows:

- Produce repeatable processes that can accurately identify the conditions set out above;
- Produce lists of actions for the appropriate business units regarding any defects found;
- Produce recommendations to improve time series data quality based on the outcomes of the project.

These high level objectives were mapped onto two project phases and translated to further low level objectives as follows:

Phase-1 Initial source data repositories for analysis

- Check and use as input previous work on a selected set of data to inform the project/study;
- Further develop the data inspection and rectification capability to:
  - Develop a pilot analysis capability based on source data;
  - Widen the geographic coverage of this capability beyond initial limits;
  - Generalise the capability to other voltage levels;
  - Automate/"productionise" the capability;
  - Carry out the initial area analysis;
  - Cross reference the different WPD analogue data management tools to check consistency and highlight issues;
  - Feedback on errors to obtain (persistent) corrections in place
  - Present a final dataset for the initial area for onward usage (analysis).
- Document the process.

Phase-2 Look at extended/enhanced Analysis capabilities

- Expand phase-1 to include e.g. power in /out analysis to provide further checks. Links to network diagram integration;
- Explore the various errors to understand these and examine ways to rectify them.

The project focussed mainly on the analogue data from the PowerOn control system as this was a priority for the Control Managers. Data was cross referenced to the Datalogger system and in a few cases metering data was used to compare to erroneous analogues to try and pinpoint the source of the problems.

3 Success Criteria

The broad criteria used to assess the success of the project were as follows:

- Accurate identification of the conditions set out in the scope;
- Successful understanding and correction of defects found;
- Improvements made to business processes based on recommendations from the project;
- Gathering of information points to help shape future strategy in terms of analogue data handling.

In more detail:

- Initial statement document defining the scope of the work, processes, methods, approach etc;
- Any tools identified and developed/obtained where necessary;
- Data processed, issues flagged, rectification programme underway;
- Method documented for future usage at other voltage levels/across further areas.

4 Details of the Work Carried Out

The project was made up of a number of distinct work elements, with bespoke data analysis tools first being developed to operate on the sample data and understand the main issues before proceeding to explore the causes of problems and find potential solutions. These activities are described below.

For more information on actual project results derived from the use of these tools see the Appendix at the end of this document.

4.1 Cross Reference Analysis – PowerOn NMC System to Datalogger Offline Data Access System

In order to look more closely at the analogue data and assess the values for errors and inconsistencies, the support teams for PowerOn and Datalogger were requested to provide a complete two year dataset (covering Jan 1 2014 to 31 Dec 2015). Additionally, a PowerOn – Datalogger X-Ref table was also made available to tie these two datasets together.

The daily datasets from PowerOn and Datalogger) were small enough (at around 10000 rows each for the SW region) to fit easily inside an Excel workbook datasheet. To facilitate the analysis, daily data from these sources was imported into multiple sheets in Excel, and VBA routines used to match between them on a line-by-line lookup using unique keys (where this was possible) to relate the two datasets to each other. This allowed the following to be done:

- Match PowerOn daily dataset to Datalogger-PowerOn X-Ref Dataset. The latter file was a cross reference between PowerOn and Datalogger and was a subset of the total Datalogger set of logger entries. In this case having an established PowerOn linkage. The matching could be done using three fields which uniquely identify each record between the 2 datasets. The fields were:
  - Site Name
  - Circuit Number
  - Unit Number
- There may be PowerOn entries not present in the Datalogger X-ref file or Datalogger X-ref file entries without a corresponding PowerOn entry. Checks were done in both directions to determine the status of the various measured entities.

- Duplicate entries were also located where two Datalogger entities matched against a single PowerOn entry. Some 20 duplicates were found in the SW region.
- An IBM SPSS based analysis showing mismatches between the units on Datalogger and the units on PowerOn was carried out and the differences tabulated. There were actually quite a significant number of instances where the units did not match, so there were cases where Amps in PowerOn were displayed on Datalogger as MW and all other combinations in between.
- Check between Datalogger/ PowerOn cross reference file and Datalogger export file. These two datasets contained a number of different entities, not least the set of derived channels which existed solely within Datalogger itself and being values which are obtained from combinations of other raw values. There were also inactive channels and also monitored entities which have existed in the past but have been deleted (or superseded at some point).
- Within the individual datasets, checks were done for all zero values, a condition which may indicate an analogue on the open side of a switch but which in error cases can be expected to reflect an invalid sensor, for example, processing links not set up correctly, not correctly commissioned or extended communications system drop.
- Within the individual datasets, checks were done for non-varying non-zero values, perhaps indicating a “stuck” or incorrectly configured sensor. To achieve this each data row was fit to a straight line using a “least squares regression” method. Gradient zero with R2 value of 1 indicates a flat straight line of perfect fit, with mean and intercept both indicating the value.
- Clearly visible in the datasets were examples of analogue values with integer (whole number) values only rather than the more usual real values (including decimal values). These examples are believed to be cases where the RTU had been configured to give this form of readout. Each case of all integer and some integer values were detected.

4.2 Data Analysis and Visualisation

Key to the rapid assessment of the data was the ability to present this graphically to an Engineer for inspection by a trained expert and including the ability to apply simple analysis functions to it for deeper understanding. By providing a quick inspection capability which did not rely on the user having to become engaged in a labour intensive process, it was possible to make a number of “quick wins” and gain a feel for the form and appearance of the data plots for the various types of analogue. This approach then paved the way for further project activities.

The visualisation and analysis program that was developed allowed the user to read in a single PowerOn Dayfile and/or multiple seasonal average files (usually four files, one for each season) which were individually created in advance using the Multifile Processor Utility Program, and then view/manipulate this data.

To create the season files, the multifile processor read in all the files placed in a target folder and averaged these on a per day basis (i.e. all Mondays in the data are summed and averaged, all Tuesdays and so on). With around 90/91 days in each season this resulted in each daily average value being formed from some 13 contributing days of that type. As the file processor utility operated on all the files deposited in the target directory it did not need to be limited only to seasons but could form monthly, annual or any other period averages.

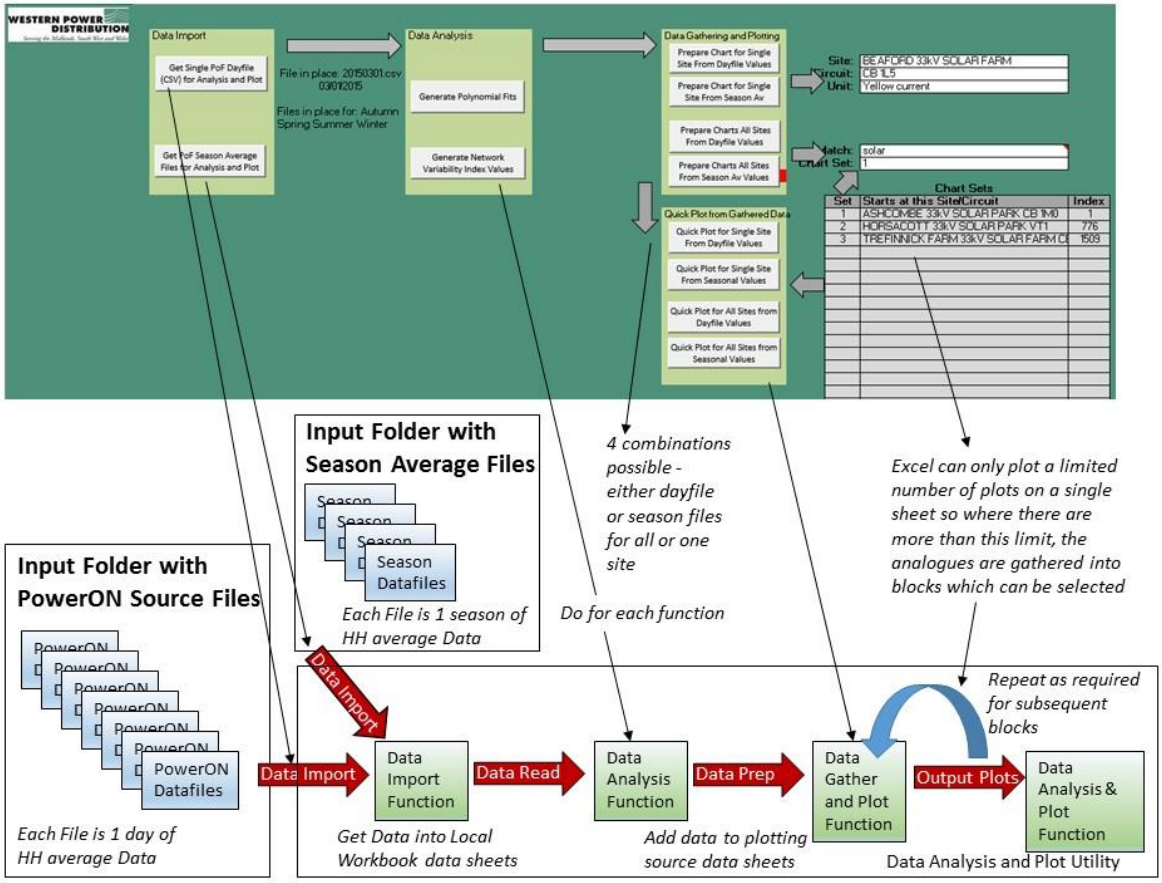

Figure 1 - Data Analysis and Plot Utility – Execution Flow

Once the data has been imported by the utility it could further be assembled for graphical plotting. It was also possible to request the generation of polynomial fit curves for all the data on the input sheets. The default polynomials are 6th order across all 48 HH values for a given day (for a single day import) or season average day and the values were appended at the end of the raw data across the row for each analogue. The data plotting function allowed the user to select whether day or season plots were required by clicking on the appropriate buttons, and for season plots which multiple season files (any/all), were to be

plotted[1]. The plots could also be requested to include fit values as well as max/min values on each day type sheet and in what form (per day or global max/min). It was also possible to specify the graph types as basic (Excel default type) or enhanced (black background).

The following option selections are available to the user in the supplementary data entry fields on sheet 12 (Program Inputs).

Parameter	Value	Comment
Chart Plotting - Max charts per sheet	128 (typically)	Excel may error if charts >128 (memory limit). 128 is the max value.
Chart Plotting - First Day to plot from left	Monday (or day)	Only works during file Data Import phase
Chart Plotting - Max / Min values by day	Yes/No	By day (Yes) or across the whole 7 days (no)
Chart Plotting - display min on graph	Yes/No	Applies to season average files only. Form of display is controlled by the selection specified above.
Chart Plotting - display max on graph	Yes/No	Applies to season average files only. Form of display is controlled by the selection specified 2 lines above.
Chart Plotting - display fitting on graph	Yes/No	If this is specified, the fit data is collected onto the charting values sheet (sheet 9) and will subsequently be plotted on any chart requests.
Chart Plotting - Autumn averages on graph	Yes/No	Applies to season average files only
Chart Plotting - Winter averages on graph	Yes/No	Applies to season average files only
Chart Plotting - Spring averages on graph	Yes/No	Applies to season average files only
Chart Plotting - Summer averages on	Yes/No	Applies to season average files only

[1] The derived polynomial coefficient values are not currently output by the program. This may be a future enhancement if it becomes desirable to derive template curves for certain analogue types.

graph		
Chart Plotting - Default or Enhanced Style	Default/ Enhanced	Default white background, Enhanced is black

Based on the user selection, the data to be plotted was assembled on a dedicated sheet into a format which was aligned to the way Excel graph plotting works and so includes a serial time point index (1 to 48 for dayfiles, 1 to 336 for the 7 day types of each season based plot). Charts were automatically generated. Use of the wildcard "*" on the match field selected all sites for plotting; otherwise any occurrence of the target text in any site name field would be selected and plotted.

Site and circuit are mandatory and act to direct the program as to what data to fetch. The program would alert if these fields were not specified. If a UNIT was not specified, all units for the Site/circuit combination were returned along with the other optional data as specified on the Program Inputs.

The "*" in the match field was a wildcard meaning all sites. Other than this special value the string entered in the match field was used to choose any site name containing this value and fetch the data for plotting. Entering Solar, Wind or BSP would collect all sites of this type (Provided a naming convention is adhered too). Alternatively it could be used to fetch and plot data for all sites called "Grantham" (the site name and the match field are both converted to upper case before the test is carried out to ensure that case is not significant).

4.3 Extended Data Analysis

Beyond simple presentation of the data to the Engineers, a capability to apply numerical analysis and engineering validation techniques to it is vital. One of the simplest checks that was envisaged during the project as being possible was to carry out was the summing of the current in and out at connected points with verification of the overall difference effectively summed to zero. However the PowerOn and Datalogger datasets had no connectivity information and the time available precluded the import and manipulation of this from a suitable system. Instead, a simpler approach was taken to sum the (yellow[2]) current on the transformers within the primaries and compare this to the sum of the current on the connected feeders of the same voltage level. The datasets included natural groupings of analogues by site so that this current summing comparison could readily be made.

The pilot program which was developed in VBA (embedded in Excel) therefore performed current summing at transformers and feeders at a given site and compared these to locate both erroneous conditions and instances of generation.

[2] Yellow current analogues are widely present in the field, having been chosen as the main preference for instrumentation. Red and Blue current analogues do exist but are much less common.

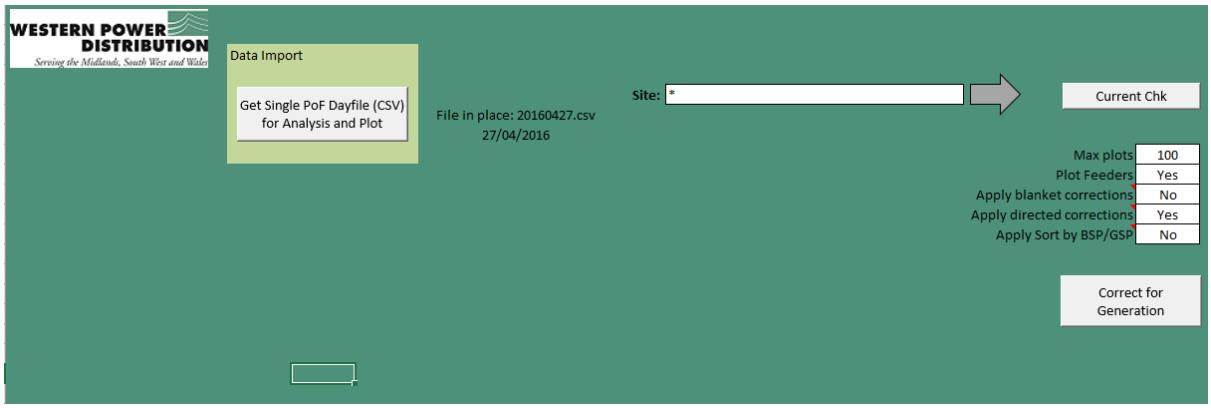

Figure 2 - Current Checker Utility - Control Sheet

If looking for instances of generation it was found to be beneficial to use a Spring or Summer weekend day already known to have a good solar response due to universally clear skies, from analysis it was found that April 18th 2015 was a good example of such a day. Other input days could be chosen for comparison with the various different analysed days to eliminate periods of zero data on some analogues.

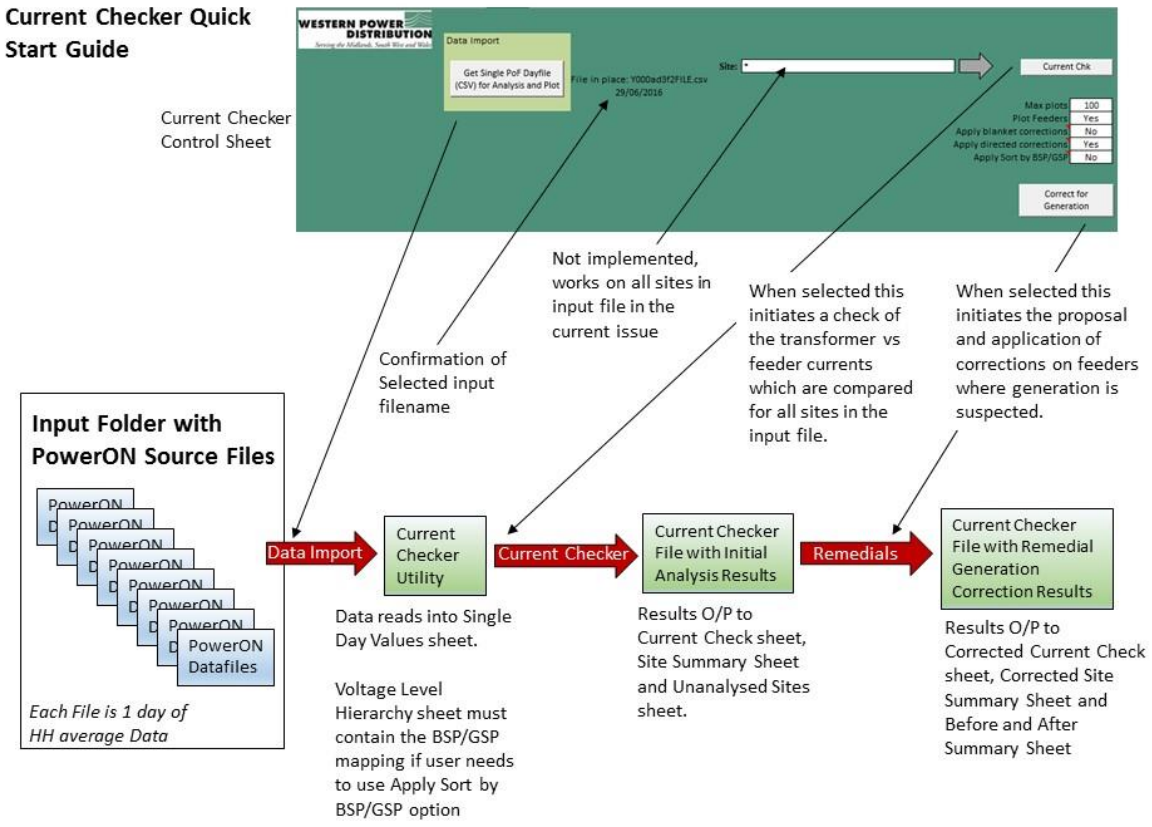

Figure 3 - Current Checker Execution Flow

When initiated, PowerOn data for a single day selected by the user is read-in and processed lexically by site using known naming rules for the different WPD regions. The presence of transformers at the site is first established, since further processing is pointless if none are present as there is nothing to compare against.

A major limitation of the program was the close dependence on the correct naming for substation objects having been observed by those setting these up and populating the PowerOn database of monitored entities. Failure to follow these naming rules (which are by necessity mirrored in the program) would render the program unable to parse the object names successfully. In such cases the comparison will almost certainly generate poor results, but the visibility of such cases in the summary and plots allows the reasons to be investigated and corrected if possible.

The naming rules are defined formally only for the SW operating region where the rules are followed quite closely. E/W Midlands and Wales actually follow their own pre-existing naming schemes. This is currently being looked at as part of a review of the findings from this project. Object names can be subject to typographical errors and other inconsistencies.

The user is informed of the assessment of each site at the analogue (component level) by the annotating of the raw input values sheet. Each analogue is by default black until analysed. Detected transformers are shown in yellow, feeders in blue and bus section/other non-summed components in red. By inspecting this assessment the user can determine how the program has treated each component and thereby assess how effective this analysis has been/identify why this course of action has been followed. On the current check sheet the attributed level of the overall site and the transformers at it are also shown (as a rounded integer). Again this allows for an assessment by the user of how the program has treated each site and component.

HEAVITREE	CB 29/13	6	Yellow current	27/04/2016	0	0
HEAVITREE	CB 29/14	6	Yellow current	27/04/2016	28	26
HEAVITREE	CB 29/15	6	Yellow current	27/04/2016	20	18
HEAVITREE	CB 29/16	6	Yellow current	27/04/2016	28	30
HEAVITREE	CB 29/17	6	Yellow current	27/04/2016	33	32
HEAVITREE	CB 29/18	6	Yellow current	27/04/2016	52	52
HEAVITREE	CB 29/19	2	kV	27/04/2016	11.02	11.02
HEAVITREE	CB 29/19	6	Yellow current	27/04/2016	130.08	122.46
HEAVITREE	CB 29/19	8	Power MVA	27/04/2016	2.48	2.34
HEAVITREE	CB 29/21	2	kV	27/04/2016	10.86	10.86
HEAVITREE	CB 29/21	6	Yellow current	27/04/2016	127.08	120.05
HEAVITREE	CB 29/21	8	Power MVA	27/04/2016	2.39	2.26
HEAVITREE	CB 29/22	6	Yellow current	27/04/2016	43	38
HEAVITREE	CB 29/23	6	Yellow current	27/04/2016	15	14
HEAVITREE	CB 29/24	6	Yellow current	27/04/2016	12	11
HEAVITREE	CB 29/25	6	Yellow current	27/04/2016	19	16

Figure 4 - Raw Data Analysis Flagging by Current Checker Utility – Annotated Input Sheet

The main analysis data output by the program is directed to a "Current Check" sheet in graphical format and to a "Site Summary" sheet in heat map format where the transformer is displayed as a HH percentage difference scaled for plot using a colour map.

By this means close correspondence is displayed in dark green, grading through lighter green, yellow, orange and then darkening red shades to poor correspondence.

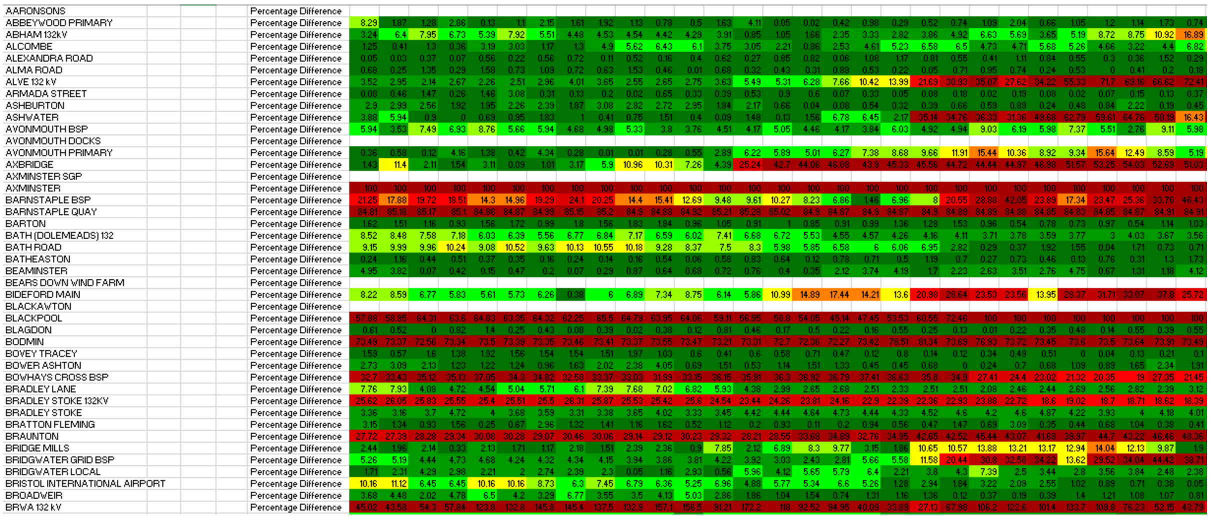

Figure 5 - Site Summary Sheet Output – Current Checker Utility

Refer to Appendix 1 – Data Analysis Results for example output results.

Sites which cannot be analysed are listed on a further output sheet. Usually such sites have one of two feeders only or have no located transformers due to various issues with object naming mentioned previously or through being substations proving a supply direct to a large consumer. Those on the unanalysed list for which the name implies some form of generation are flagged in grey.

Generation was found to be readily identifiable on the heat map site summary plots by the presence of blocks of dark red on an otherwise green line, though it may not always be so simple. Red bands in the middle of the day may be indicative of solar generation presence on an attached feeder. The current check sheet has plots of the analogue values and feeders can be selected for plotting and allow the individual feeders to be checked for involvement in generation.

Two options were provided on the control sheet to allow the user to apply corrections for possible generation. The options are "Apply blanket corrections" and "Apply directed corrections". In blanket correction mode any feeders which have a name indicative of generation have their sense flipped and subtracted from the feeder total rather than being added in as us usual. This is a simplistic approach.

In directed correction mode each transformer current difference value (from the 48 for the site) is inspected and if this exceeds a threshold percentage then the value of the feeder (yellow current value) closest to the difference value has its sense flipped. The colour of any flipped feeder is flagged by setting its value in that HH slot to have a blue background in the revised current check sheet output and the totals are recomputed and flagged in grey.

Looking along the values for a given site soon reveals by inspection which feeder is the most likely to be the correcting feeder. Correcting feeders are flagged on the current check sheet along with a count being output against each feeder of how many times its value has been flipped. There is also a revised site summary sheet reflecting all the sites in the revised/corrected analysis and a final before and after site summary presentation allowing ready assimilation of the results.

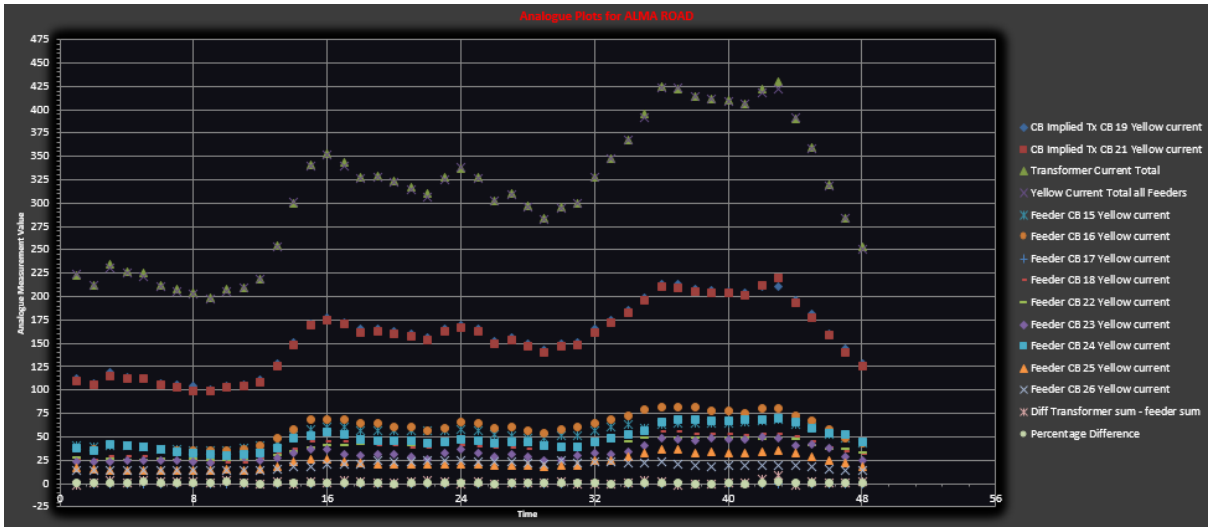

Figure 6 - Site Graph from Current Check Sheet

Figure 6 above shows the graph of the identified elements for a particular site – Feeders, transformers and summed values. It is readily seen that the transformer and feeder yellow current totals are in very close correspondence (topmost points – green triangle and purple cross), the two detected transformers (red square and blue dot in the lower middle of the graphing area) are also in close agreement with each other so are sharing load). The feeders are plotted near the bottom having relatively low yellow current values while the white dots (percentage difference between feeder current and transformer current totals) are very close to zero at all points.

4.4 Stakeholder Engagement

The analogue data consumption stakeholders were identified at project start up as follows:

Stakeholder	Interest
Control Managers	Effective Network Management
Control Engineers	Network monitoring, situational awareness and network control.
Planning Engineers	Network assessment, planning, general investigative work.

As part of closure, other DNOs will be consulted.

4.5 Pilot Rectification Programme

During the commencement of the project, Management were appraised of the initial findings and based on the overall data quality analysis results the project was able to compile a snagging list for a sample area and initiate rectification in the field to assess the size of the overall task were it to be scaled to cover the full analogue set. An active ANM enabled area was chosen together with a number of key sites exhibiting signs of reverse power flows where only current analogues exist.

5 The Outcomes of the Project

In terms of the scale of data quality issues and identifying other key problems, the main project conclusions are as follows:

Findings:

- 13.8% of all analogues in the WPD South-West licence area are only recording 0 values. (20.7% companywide). Many of these may be valid open circuit values, however some will reflect incorrect values;
- 1% of all existing South-West PowerOn data points are not available for planners in Datalogger. Compared to 12% in Wales and 36% in the Midlands;
- 63% of all new solar sites across the company have not had their analogues commissioned correctly;
- Policy document ST:OC12B/3 regarding the numbering and labelling of substation apparatus is not being consistently applied across all four licence areas;
- In the Wales and Midlands regions, the numbering and labelling of substation apparatus follows different rules and object naming rules need to be more rigorously enforced in the future;
- When defining the analogues to the system, data entry is free format and not effectively validated to disallow obviously incorrect entries. The systems would be much improved by enforcing some form of validation at entry time;

- 204 circuits within the South-West area (326 companywide) experience reverse flows which are not apparent from the existing analogue values.

Actions:

- The production of a 'Monitoring System Handbook' would allow for a better understanding of the end-to-end compound data system, and give and understanding of what resolutions are obtained for the various areas of uncertainty identified in this report;
- It is recommended that a 'Time Series Data' policy is produced, clearly outlining Data owners, their responsibilities and rectification process;
- More rigorous housekeeping in the PowerOn and Datalogger systems is required based on the initial outputs from this project. It is the case that there are data quality reports already being routinely produced but these are not rigorously acted upon;
- Having identified several classes of errors in the data, a number of these have been successfully rectified by engineers in a pilot rectification programme and the causes of errors and effort required to correct these has been analysed. This initial pilot rectification programmes should be continued to extend the remedial actions already taken to correct the issues identified across the whole analogue set and also where possible to remove the sources of such errors (for example when setting up analogues – to utilise drop-downs for restricted choices from lists rather than allowing user free-form entry (which is a particularly error prone way of defining objects, leading to many opportunities for local divergences in style, or errors);
- A further project phase (Part 2) may be initiated to investigate Data Management tools (see below).

Further Developments:

- It has been demonstrated in pilot that it is possible to visualise, analyse and model the analogue data using techniques such as automated bulk data plotting, curve fitting (by regression analysis), current summing, averaging (by season for example), template "typical response" creation and other statistical methods. These techniques may be readily deployed to give Engineers a better insight and may also be extended in the future by correlation with other datasets currently held in isolated, unconnected databases or held externally (to WPD) by other third parties such as the Met Office;
- All highlighted issues with individual analogues to be exported to a new rectification tracking and reporting tool (already under development);
- Work to define a unique object identifier to be used across all systems to avoid reliance on string matching when looking to combine data from multiple systems.

6 Performance against Project Aims, Objectives and Success Criteria

The analytics deployed have followed two main strands:

IBM were used in a short focussed engagement and have used their SPSS analytics tool to look at both samples of the data and block data extracts, and using bespoke data analysis tools developed in Visual Basic, some of which could be compared to the IBM results.

Initially looking just at the SW operating region of WPD, the scope was expanded to include time series analogue data from all of WPD's regions once the tools were developed. Large amounts of data were processed to carry out the following: Cross reference different datasets to pinpoint inconsistencies and generate reports for further corrective action; develop an automated block data plotting and curve fitting function based on half hour average values allowing fast review by engineers. By plotting all analogues it was clear where (for example in cases of generation) the directional sense of analogues was incorrectly set.

Further visualisation methods were also explored in particular heat mapping to illustrate both missing data, cases of incorrect values, data spikes and to also illustrate actual "real" trends. An innovative tool was also developed in pilot to check yellow current values at sites with transformers, to compare summed current at the transformers and along the feeders and reconcile the two. Where a missing analogue could be inferred (for example from MVA values and the prevailing kV) these were included and the analysis expanded. Application of correction strategies in possible cases of generation were explored where the current differences were in excess of a threshold, by (for example) flipping the direction/sense of a suitable candidate feeder. Taken together all these tools allowed a priority list of analogue rectifications to be prepared and handed to BAU for implementation - a programme of which is ongoing with high level management and control room support.

7 Required modifications to the planned approach during the course of the project

It was relatively easy to expand the scope of the investigation in a cost effective manner to cover all WPD operating regions once the automated tools had been perfected, so this was done in order to make the project more effective at no extra cost. The key point here being early tool development. Rules were added as necessary where there were regional variants to be handled, but otherwise all that was required to be done was to secure the necessary set of exported day files from the relevant PowerOn system and then execute RUNs of the tools against these. Object naming was found to be an area of unexpected non-compliance so the rules were reviewed and issues documented for future action. In other respects the work proceeded largely as planned with the specific utility/tool suite developed in an agile

fashion to suit the perceived needs of the analysis. This aspect was not planned in detail prior to the project start. The project also explored more in depth data analysis techniques given that data was available to allow this to be done.

The number of issues found was larger than initially foreseen so Senior Stakeholders were appraised of the situation in order to gain sponsorship for a rectification programme which was initiated to determine the size of the next phase remedial programme and gather details of the means for the resolution of the various issues found.

8 Significant variance in expected costs and benefits

The project remained within cost throughout; the initial IBM consultation was not extended beyond presentation of their initial findings which had served to effectively validate the other work stream. Tools that were developed by the project were provided using the standard Microsoft office IT suite of programs, specifically MS Excel with its embedded VBA (Visual Basic for Applications) capability. This made the tools highly portable and capable of being distributed to new users on request as well as obviating the need for additional software licences, all of which enabled us to keep costs down.

Benefits were in line with our initial expectation, with the success of the analysis work feeding through not only to the Half Hourly average values used for offline inspection, but also informing the PowerOn Control system concerning redundant or unnecessary analogues not needed by the real-time Control systems themselves. With performance always a concern in this area, the presence of additional processing associated with redundant analogues is to be avoided if at all possible.

9 Lessons Learnt for Future Projects

The main recommendations are already being actioned in an analogue priority items rectification programme. Our Policy is also being reviewed in respect of object naming, the in-field implementation of which has been identified as problematic. The resulting actions are expected to facilitate data clean-up as well as being of use to *Control* as unused objects are being removed from the monitoring database in a bid to improve performance in control systems where this is a concern.

The curve fitting tools developed by the project were also investigated in respect of derivation of "template responses" for certain analogue types, e.g. Solar response by season. More work in this area may yield further useful capabilities in the area of capacity determination, visualisation and management. It is also the case that this project benefited from a number of data analysis and visualisation techniques that had been developed and used on the WPD FALCON Project, in particular the use of “Heat Maps” to visualise large volumes of data. This happened readily because of staff continuity carrying over from FALCON, but where that is not the case the recommendation would be for start-up projects

to check previous project summaries for their recommendations, conclusions and tools (i.e. conduct a *literature search*!)

Microsoft Excel has proved to be a very powerful platform for tool development and rapid pilot utility demonstration pieces. When enabled with MS VBA (Visual BASIC for Applications) the combination of easy coding environment, output framework (the spreadsheets), linkage to other database systems such as MS Access and SQL Server plus the fact that this is all licenced widely for most users, without any additional expense, means that developments can be very rapid and effective.

Another more advanced system to supplement and replace the pilot tools developed in the work is currently being scoped and specified with a view to rolling this out to Engineers for better access and facilities relating to data management, visualisation and analysis and in particular the rectification of erroneous data. Additional analytical tools may become available through the deployment of commercially available data management systems once the requirements have been defined following the pilot investigation work. Pilot deployment to a new data management production system is being investigated as a follow-on activity.

The main problems discovered were in the data itself which tripped up the analysis or made it more complex. Of particular note was the impact of object names on the ability of the pilot tools to parse data files and conduct reliable inference analysis in the absence of descriptive attributes relating to the data itself (this aspect may be resolved in the future with the possibility of linkage of the next generations tools to a projected INM database). In the current pilot tools, the data files only refer to object names, not their type, so the type must be inferred from the name. Errors in the data therefore make this analysis very much more difficult than it perhaps needs to be.

Better than reliance on string matching however, particularly when looking at data from multiple source systems, would be to deploy a UNIQUE IDENTIFIER for objects across all systems. Work is taking place on the WPD INM project and may build on early work in this area to define a master “Network Reference ID” across systems. At the present time this is a CROWN (asset management system) attribute which is being used for data exported to the WPD public facing website. This is a work in progress.

By using visualisation tools, the data analysis was readily demonstrated to stakeholders and peers with similar data access requirements. It was easy to get stakeholder approval for the rectification programme work once the impact of the pinpointed issues was made clear. The results were very clear and indisputable so this was an achievable objective. The use of IBM (and their SPSS tool) early in the analysis also provided a useful cross check of the bespoke tools as many of the conclusions corresponded.

10 Planned Implementation

Not only are the pilot tools developed during the project being maintained and further developed for immediate ongoing use by Engineers, a production system based on commercially available offerings/systems is already being scoped based on the findings of the pilot tools deployed and used in this project. This system may in future be integrated with data in other ongoing parallel activities such as the Integrated Network Model (INM) initiative which is capitalising on the Authorised Network Model (ANM) of FALCON, another WPD LCNF project which ended in 2015.

The FALCON Project Authorised Network Model merged data from the various island databases of network data in WPD, including sources such as PowerOn, Crown and EMU, and the new INM will do the same for a larger pilot area in Cornwall. An export from this database would be able to provide much more detailed asset and connectivity information.

11 Facilitate Replication

Having now carried out the main investigative activity and set in train remedial actions in respect of the data issues identified, it is unlikely that replication of this project within WPD would be necessary or desirable unless the remedial work programme ceased before completion and/or an effective forward corrective process for data issues was not put in place. As the tools used will be retained, these can simply be deployed and used again as required and to this end a User handbook and quick start guides have been prepared.

For the benefit of other DNOs, the findings and conclusions of this project and the description of the approach taken and tools developed contained in this document should be of use in either starting up a similar project or informing one already under way. WPD intends to conduct a dissemination event to make this information available in a ready manner to interested parties.

11.1 Knowledge Required

The main knowledge areas for the project are:

- Familiarity with the existing analogue management systems and field deployed equipment;
- Data analysis techniques;
- IT aspects of data processing (including familiarity with the tools developed by the Project) and;
- Provision of engineering insight to guide development of tools and interpret results.

For other DNOs, while their systems, tools and methods may differ from those specific aspects within WPD, the overall concepts explored and conclusions reached by this project will be very similar.

11.2 Products/Services Required

These are the key elements to allow for successful replication of this project by another DNO:

1. Stakeholder support;
2. Provision of input data for analysis;
3. WPD assist/advice with preparatory work.

12 Contact

Further details on replicating the project can be made available from the following points of contact:

Future Networks Team
Western Power Distribution,
Pegasus Business Park,
Herald Way,
Castle Donington,
Derbyshire
DE74 2TU
Email: wpdinnovation@westernpower.co.uk

Reference Documents

1. IBM Initial Data Analytics Engagement Final Report. Final report v1.0.pptx. Jonathan Batson, Kelly Freeman, Erwin Frank-Schultz. March 2016 Version
2. Company Directive STANDARD TECHNIQUE: TP6F, Power Measurement Conventions, February 2016
3. BSEN 60044-1:1999 Instrument transformers – Part 1: Current transformers
4. Company Directive STANDARD TECHNIQUE: OC12B/3, Numbering and Labelling Switchgear and other Apparatus within Substations, November 2004
5. Company Directive STANDARD TECHNIQUE: GE23/1, Relating to Data Assurance of Regulatory Submissions, October 2015
6. Analogue Data Quality Analysis Report – Cross Reference Report, July 2016
7. Analogue Data Quality Analysis Report – Heat Map Report, July 2016

8. Analogue Data Quality Analysis Report – Seasonal Analysis Report, July 2016

Appendix 1 – Data Analysis Results

This section includes some sample results from the analysis runs of the various utilities and tools developed during the project. These are provided as examples. For further more in depth information and examples see the individual Project reports in Refs [6 - 8].

Data Plotting

Two main types are available – based on being single day or seasonal averages. Examples are shown below.

The first three plots are seasonal averages for solar parks (specific site selected and anonymised for the illustration) and show available analogues (Summer only selected) along with polynomial fit and the derived global maximum value for each analogue (taken from all inputs, horizontal bars).

Each day is graphed along the plot with Monday on the left (this is a program selection). The key shows colour/tick mark correspondence. In the second plot for Solar Park 2 the presence of a potential directional error is apparent. For the Power MW analogue which is showing a positive peak in phase with the max current, such an instance would be flagged as an error for further investigation as a positive export is unusual, but may be correct dependent on the position of the local measurement CTs.

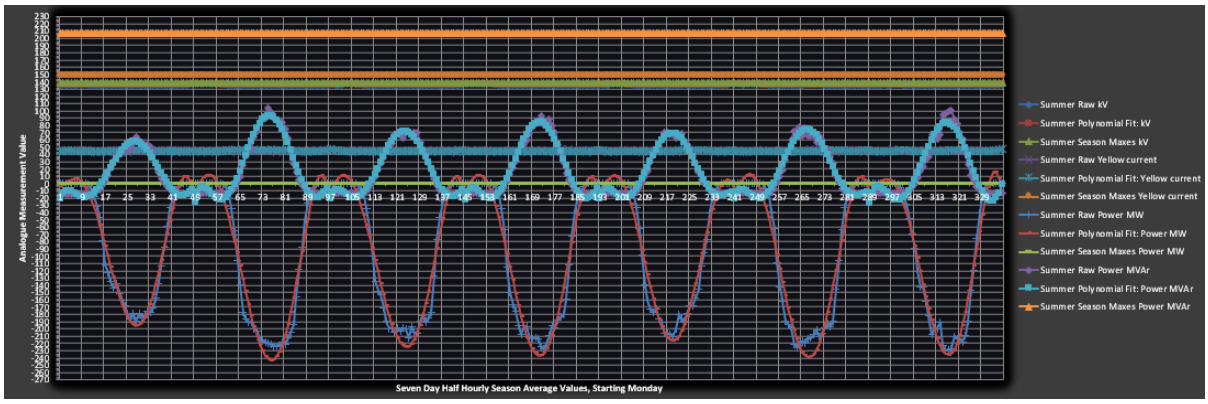

Figure 7 - Plot Utility - Solar Park 1 (132kv)

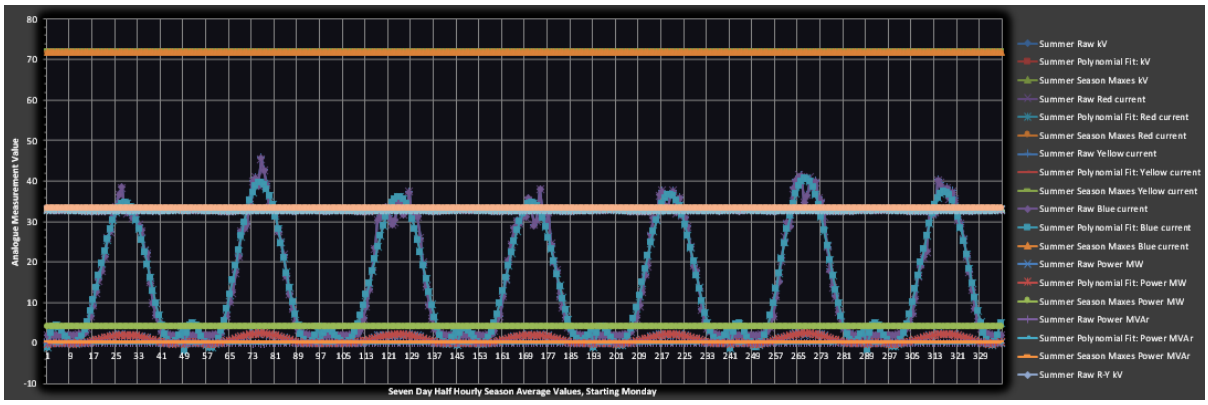

Figure 8 - Plot Utility Solar Park 2 (33kV)

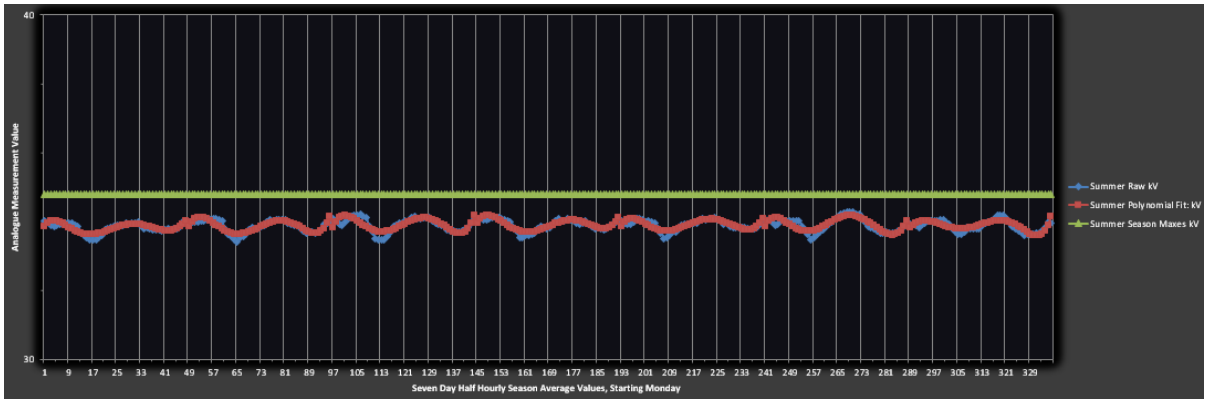

Figure 9 - Plot Utility – Solar Park 3 Voltage Plot (33kV)

Data Fitting (Regression Analysis)

The plot below is a classic example of a single solar farm analogue plot on a day chosen for a good solar response. The 18th April 2015 was a very sunny day in Spring, so the response is typical, corresponding closely to the ideal. Yellow current is shown in blue points and line, while Power in MW is green and Power MVAr is plotted in cyan. The curve fits (polynomial of order 6) to these raw values and are in red, purple and orange respectively. In the cases of MW and MVAr the fit lines are so good that the raw data is almost invisible plotted underneath. For the yellow current raw (blue) and fit (red) the points of departure between the two sets of values are visible, but again the correspondence is very good except in the wings where the solar response is very low (night conditions).

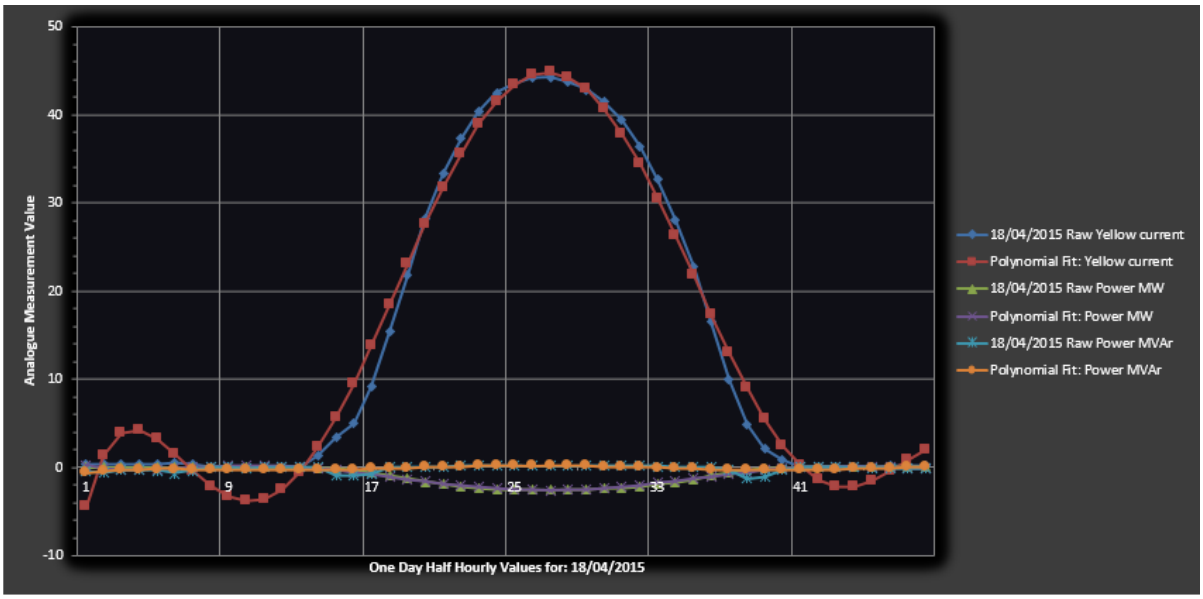

Figure 10 – Solar Park 4 Example Single Site Plot Output with Data Fitting (Polynomial Curves)

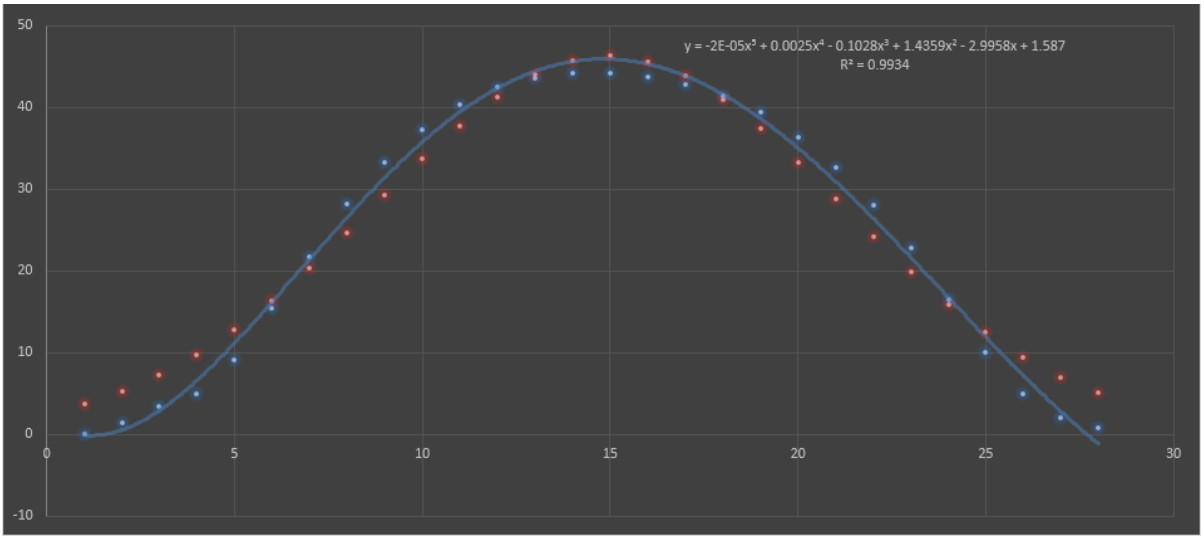

Figure 11 - Solar Park 4 Raw Data with Polynomial & Gaussian Fit

In the above plot, the circuit CB 1M0 raw yellow current for Solar Park 4 yellow current response data (blue marker points) is fit by a Gaussian regression (red markers) and a fifth order polynomial (blue line), omitting the leading and trailing wings. The polynomial gives the better fit (R2 value of 0.9934, very close to ideal value of 1) particularly at the lower values of current early and late in the day. The Gaussian Mean point μ is readily determined and pinpoints the point of peak response at local noon.

Seasonal Averaging

Some typical seasonal averaging plots are given below. When the Analysis and Plot utility executes, plots may be prepared for any of the sites in the input files covered by the period of the analysis. These examples are for solar farms for which both the diurnal form of the plot is clear (bell curve) as well as the seasonal variation (higher amplitude in Spring and Summer, lower in Winter and Autumn). For more examples refer to Ref [8].

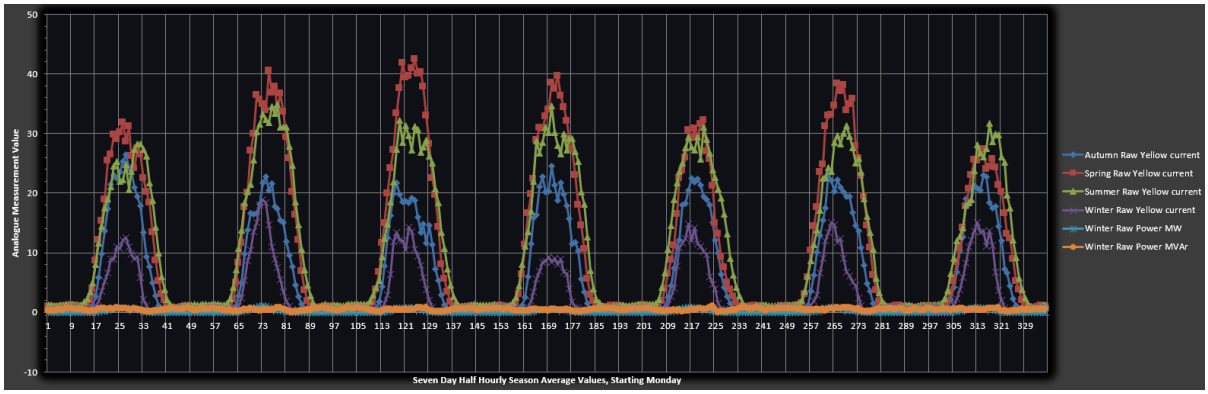

Figure 12 - Solar Site 5 - 4 Season Averages, Raw Data Only

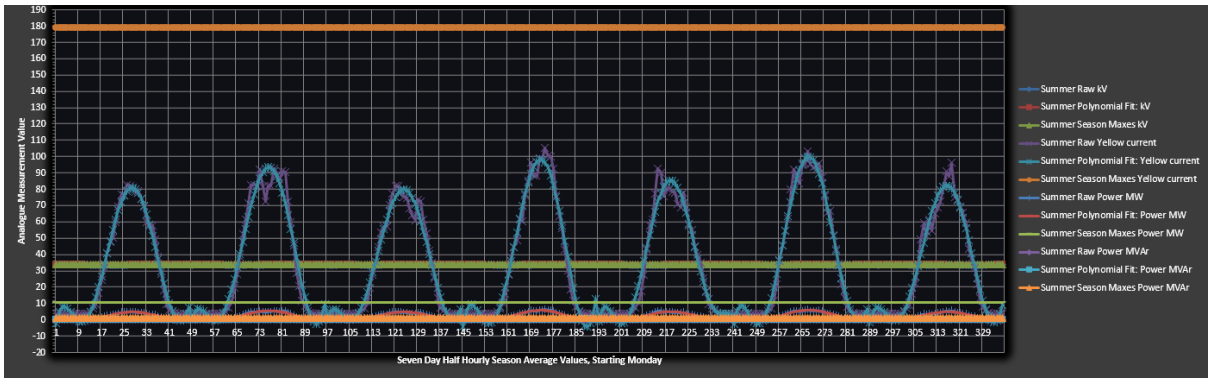

Figure 13 - Example Season Plot Showing Polynomial Fits and Maxima

Capacity Factor/Network Variability Index Determination

The following diagram shows an example colourised table of computed Network Variability Index values (a measure of network capacity). This is not strictly a heat map as the values in each row are not (in general) related to each other, but they are colourised to highlight spike values.

Site	Circuit	Unit	Autumn Monday Network Variability Index	Spring Monday Network Variability Index	Summer Monday Network Variability Index	Winter Monday Network Variability Index	Autumn Tuesday Network Variability Index	Spring Tuesday Network Variability Index	Summer Tuesday Network Variability Index	Winter Tuesday Network Variability Index	Autumn Wednesday Network Variability Index	Spring Wednesday Network Variability Index	Summer Wednesday Network Variability Index	Winter Wednesday Network Variability Index	Autumn Thursday Network Variability Index	Spring Thursday Network Variability Index	Summer Thursday Network Variability Index	Winter Thursday Network Variability Index	Autumn Friday Network Variability Index	Spring Friday Network Variability Index	Summer Friday Network Variability Index	Winter Friday Network Variability Index
ABBEYWOOD PRIMAF	CB 5	Yellow current	1.66	3.00	1.56	2.83	1.33	2.49	1.36	2.40	1.71	3.34	1.66	2.83	1.41	2.67	1.34	2.22	1.43	2.64	1.40	2.27
ADELAIDE ROAD	CB 22-New	Yellow current	1.25	1.23	1.22	1.11	1.23	1.18	1.21	1.11	1.23	1.19	1.20	1.11	1.25	1.19	1.20	1.15	1.21	1.17	1.17	1.14
ADELAIDE ROAD	CB 23-New	Yellow current	1.25	1.25	1.20	1.16	1.24	1.26	1.18	1.18	1.32	1.32	1.24	1.24	1.25	1.23	1.18	1.16	1.22	1.22	1.16	1.23
ADELAIDE ROAD	CB 24-New	Yellow current	1.36	1.43	1.79	1.17	1.34	1.38	1.72	1.12	1.39	1.41	1.76	1.13	1.43	1.47	1.87	1.17	1.40	1.40	1.81	1.16
ADELAIDE ROAD	CB 25-New	Yellow current	1.98	1.98	2.12	1.55	1.90	1.61	1.77	1.37	1.87	1.63	1.71	1.32	3.24	3.47	3.65	2.36	1.71	1.50	1.59	1.30
ALCOMBE	CB 119	Yellow current	1.78	1.49	1.91	1.59	2.20	1.68	1.91	1.97	2.07	1.64	1.75	1.96	2.26	1.81	1.96	1.87	1.79	1.73	2.22	1.41
ALCOMBE	CB 132	Yellow current	1.78	1.50	1.93	1.59	2.26	1.73	1.91	1.96	2.05	1.63	1.76	1.94	2.26	1.81	1.97	1.86	1.79	1.73	2.22	1.40
ALCOMBE	CB 133	Yellow current	2.67	3.69	2.67	3.83	2.38	3.34	2.49	3.47	2.55	3.40	2.71	3.20	2.63	3.26	2.78	3.46	3.21	3.80	2.89	3.35
ALCOMBE	CB 37	Yellow current	1.49	1.56	1.70	1.31	1.48	1.54	1.56	1.27	1.51	1.57	1.65	1.27	1.65	1.77	1.86	1.36	1.53	1.64	1.32	1.23
ALCOMBE	CB 37	Power MVA	1.47	1.54	1.68	1.30	1.46	1.52	1.54	1.26	1.48	1.54	1.62	1.25	1.63	1.74	1.82	1.37	1.53	1.64	1.32	1.29
ALCOMBE	CB 38	Yellow current	1.76	1.88	2.17	1.52	1.95	2.12	2.11	1.71	1.89	2.01	2.23	1.60	1.54	1.71	1.91	1.38	1.35	1.40	1.67	1.15
ALCOMBE	CB 39	Yellow current	1.30	1.41	1.76	1.08	1.38	1.53	1.86	1.11	1.33	1.43	1.77	1.08	1.37	1.48	1.85	1.10	1.36	1.45	1.81	1.14
ALCOMBE	CB 40	Yellow current	1.28	1.25	1.29	1.11	1.30	1.27	1.33	1.14	1.23	1.19	1.22	1.08	1.28	1.23	1.30	1.14	1.26	1.22	1.27	1.13
ALCOMBE	CB36	Yellow current	1.51	1.46	1.62	1.36	1.39	1.37	1.42	1.23	1.47	1.43	1.48	1.25	1.61	1.60	1.66	1.35	1.62	1.62	1.81	1.36
ALCOMBE	CB36	Power MVA	1.51	1.45	1.61	1.36	1.40	1.38	1.43	1.24	1.45	1.41	1.45	1.24	1.53	1.58	1.64	1.34	1.61	1.62	1.81	1.36
ALEXANDRA ROAD	CB 18/14	Yellow current	0.00	13.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ALEXANDRA ROAD	CB 18/15	Yellow current	1.17	1.19	1.21	1.05	1.15	1.15	1.16	1.04	1.18	1.19	1.20	1.07	1.20	1.15	1.19	1.08	1.16	1.17	1.19	1.10
ALEXANDRA ROAD	CB 18/16	Yellow current	1.32	1.36	1.26	1.14	1.30	1.18	1.23	1.13	1.24	1.17	1.16	1.10	6.76	6.68	6.64	5.07	1.18	1.15	1.14	1.13
ALEXANDRA ROAD	CB 18/17	Yellow current	2.25	2.79	3.71	2.07	2.34	2.75	3.55	2.11	2.47	2.70	3.68	2.02	9.96	10.81	15.20	5.84	2.21	2.54	3.73	2.02
ALEXANDRA ROAD	CB 18/18	Yellow current	1.29	1.39	1.61	1.08	1.40	1.42	1.68	1.13	6.75	7.08	8.14	4.39	1.42	1.45	1.72	1.14	1.34	1.35	1.64	1.13
ALEXANDRA ROAD	CB 18/19	Yellow current	1.28	1.38	1.52	1.15	1.32	1.35	1.53	1.15	1.24	1.29	1.42	1.08	1.30	1.38	2.22	1.70	1.73	1.78	2.03	1.58
ALEXANDRA ROAD	CB 18/19	Power MVA	1.28	1.38	1.52	1.16	1.31	1.34	1.52	1.15	1.24	1.28	1.41	1.07	1.88	1.96	2.20	1.69	1.71	1.75	2.00	1.56
ALEXANDRA ROAD	CB 18/21	Yellow current	1.55	1.65	1.78	1.30	5.05	5.46	6.23	4.73	1.63	1.55	1.80	1.04	1.58	1.48	1.84	1.32	2.02	2.04	2.31	2.61
ALEXANDRA ROAD	CB 18/22	Yellow current	1.90	1.83	2.12	1.50	6.09	5.75	6.78	4.30	1.39	1.38	1.60	1.09	1.38	1.38	1.63	1.13	1.45	1.33	1.60	1.13
ALEXANDRA ROAD	CB 18/23	Yellow current	1.36	1.44	1.81	1.22	8.60	9.38	11.31	6.43	1.40	1.53	1.87	1.19	1.43	1.56	1.92	1.22	1.44	1.46	1.90	1.23
ALEXANDRA ROAD	CB 18/24	Yellow current	4.74	5.29	6.48	3.03	4.81	5.39	6.72	3.09	27.08	30.60	38.41	10.38	1.55	1.79	2.17	1.21	1.43	1.63	2.03	1.14
ALEXANDRA ROAD	CB 18/25	Yellow current	1.27	1.38	1.55	1.08	1.30	1.39	1.59	1.11	15.32	16.54	18.33	7.56	1.28	1.39	1.61	1.10	1.30	1.32	1.56	1.10
ALEXANDRA ROAD	CB 18/26	Yellow current	1.29	1.36	1.50	1.16	1.31	1.34	1.51	1.16	1.24	1.27	1.39	1.08	1.96	2.05	2.28	1.77	1.23	1.24	1.42	1.13
ALEXANDRA ROAD	CB 18/26	Power MVA	1.29	1.36	1.50	1.16	1.30	1.33	1.49	1.15	1.24	1.27	1.39	1.08	1.97	2.05	2.28	1.78	1.24	1.25	1.43	1.13

Current Summing at Primaries

The following plot is a section from a current check output (all site summary). Each site is a new line in the output. Dark green values show excellent correspondence between feeder current and transformer yellow current totals, while lighter green, yellow, orange and red values show increasing divergence in these totals.

AARONSONS			Percentage Difference																												
ABBEYWOOD PRIMARY			Percentage Difference	8.29	1.57	1.35	2.96	6.19	1.1	2.25	1.61	1.92	1.13	0.78	0.5	1.63	4.11	0.05	0.02	0.42	0.39	0.29	0.52	0.74	1.08	2.04	0.66	1.65	1.2	1.14	1.21	0.78
ABHAM 132kV			Percentage Difference	3.26	6.4	7.95	6.73	5.39	7.92	5.51	4.49	4.53	4.54	4.42	4.29	3.91	0.85	1.05	1.66	2.25	3.33	2.82	3.66	4.92	6.63	5.63	3.65	5.19	8.72	8.75	10.92	16.89
ALCOMBE			Percentage Difference	1.25	0.41	2.13	0.26	3.19	2.03	1.17	1.3	4.9	5.62	6.43	6.1	3.75	3.05	2.21	0.86	2.53	4.61	5.23	6.59	6.5	4.73	4.71	5.68	5.26	4.66	3.22	4.4	6.22
ALEXANDRA ROAD			Percentage Difference	0.05	0.03	0.37	0.07	0.56	0.22	0.56	0.72	0.11	0.52	0.16	0.4	0.62	0.27	0.85	0.82	0.06	1.08	1.17	0.81	0.55	0.41	1.11	0.84	0.55	0.3	0.36	1.52	0.29
ALMA ROAD			Percentage Difference	0.68	0.25	1.35	0.29	1.58	0.73	1.09	0.72	0.63	1.53	0.46	0.01	0.68	0.32	0.43	0.31	0.88	0.53	0.22	0.05	0.71	0.95	0.74	0.24	0.53	0	0.41	0.2	0.19
ALVE 132 kV			Percentage Difference	3.52	2.95	2.14	2.67	2.26	2.51	2.96	4.01	3.65	2.55	2.65	2.75	3.63	5.49	5.31	6.28	7.66	10.42	13.99	21.69	30.93	35.07	27.62	34.22	55.35	71.7	69.16	66.62	72.41
ARMADA STREET			Percentage Difference	0.08	0.46	1.47	0.26	1.46	3.08	0.31	0.13	0.2	0.02	0.65	0.33	0.39	0.52	0.9	0.6	0.07	0.33	0.05	0.08	0.18	0.02	0.19	0.06	0.02	0.07	0.35	0.13	0.37
ASHBURTON			Percentage Difference	2.8	2.39	2.56	1.92	1.85	2.26	2.39	1.87	3.09	2.82	2.72	2.95	1.84	2.17	0.66	0.04	0.09	0.54	0.32	0.29	0.66	0.59	0.09	0.24	0.48	0.84	0.22	0.19	0.45
ASHWATER			Percentage Difference	3.88	5.94	0.5	0	0.63	0.95	1.83	1	0.41	0.75	1.51	0.4	0.09	1.48	0.13	1.56	6.78	6.45	2.17	35.14	54.76	36.33	31.36	49.68	62.29	59.61	64.76	50.18	16.43
AVONMOUTH BSP			Percentage Difference	5.94	3.53	7.49	6.93	8.76	5.66	5.94	4.68	4.36	5.33	3.8	3.76	4.51	4.17	5.05	4.46	4.17	3.84	6.03	4.32	4.34	9.03	6.19	5.98	7.37	5.51	2.76	9.11	5.99
AVONMOUTH DOCKS			Percentage Difference																												
AVONMOUTH PRIMARY			Percentage Difference	0.36	0.59	0.12	4.16	1.38	0.42	4.34	0.28	0.01	0.01	0.29	0.55	2.89	6.22	5.89	5.01	6.27	7.38	8.68	9.66	11.91	15.44	10.36	8.92	9.34	15.64	12.49	8.59	5.19
AXBRIDGE			Percentage Difference	1.44	11.4	2.11	1.54	3.11	0.09	1.01	3.17	5.9	10.96	10.31	7.26	4.39	25.24	42.7	44.06	46.08	43.9	45.33	45.56	44.72	44.44	44.97	46.98	51.57	53.25	54.03	52.69	51.03
AXMINSTER SGP			Percentage Difference																												
AXMINSTER			Percentage Difference	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
BARNSTAPLE BSP			Percentage Difference	21.25	17.88	19.72	18.51	14.3	14.96	19.29	24.1	20.25	14.4	15.41	12.69	9.48	9.61	10.27	8.23	6.66	1.46	6.96	8	20.55	28.08	42.05	23.89	17.34	23.47	25.36	33.76	46.43
BARNSTAPLE QUAY			Percentage Difference	84.91	85.18	85.17	85.1	84.86	84.87	84.93	85.15	85.2	84.9	84.88	84.92	85.21	85.28	85.02	84.9	84.87	84.9	84.67	84.9	84.89	84.89	84.88	84.85	84.83	84.85	84.87	84.91	84.91
BARTON			Percentage Difference	1.62	1.51	1.16	0.93	1.56	1.72	0.89	1.8	1.56	1.83	1.84	0.96	1.05	0.91	1	0.85	0.91	0.99	1.36	1.28	1.53	0.96	0.54	0.78	0.73	0.97	0.54	1.14	1.03
BATH (DOLEMEADS) 132			Percentage Difference	8.52	8.48	7.58	7.18	6.03	6.39	5.56	6.77	6.84	7.17	6.59	6.02	7.41	6.68	6.72	5.53	4.55	4.57	4.26	4.16	4.11	3.71	3.78	3.59	3.77	3	4.03	3.67	3.56
BATH ROAD			Percentage Difference	9.15	9.99	9.96	10.24	9.08	10.52	9.63	10.13	10.55	10.18	9.28	8.37	7.5	8.3	5.98	5.85	6.58	6	6.06	6.95	2.82	0.29	0.37	1.92	1.55	0.04	1.71	0.73	0.71
BATHEASTON			Percentage Difference	0.34	0.5	0.14	0.41	0.37	0.29	0.16	0.24	0.16	0.24	0.06	0.59	0.82	0.64	0.12	0.72	0.71	0.5	1.19	0.7	0.27	0.73	0.46	0.13	0.76	0.31	1.3	1.73	
BEAMINSTER			Percentage Difference	4.35	3.82	0.07	0.42	0.15	0.47	0.2	0.07	0.29	0.87	0.64	0.68	0.72	0.76	0.4	0.25	2.12	0.74	4.19	1.7	2.23	2.63	3.51	2.78	4.75	0.67	1.21	1.19	4.12
BEARS DOWN WIND FARM			Percentage Difference																												
BIDEFORD MAIN			Percentage Difference	8.22	8.59	6.77	5.83	5.61	5.73	6.25	0.38	6	6.89	7.34	8.75	6.14	5.86	10.99	14.89	17.44	14.21	13.6	20.98	28.64	23.53	23.56	13.95	28.37	31.71	31.07	37.6	25.72
BLACKAWTON			Percentage Difference																												
BLACKPOOL			Percentage Difference	57.88	58.95	64.31	63.6	64.83	63.35	64.32	62.25	65.5	64.79	63.95	64.06	59.11	56.95	59.8	54.05	45.14	47.45	53.53	50.55	72.46	100	100	100	100	100	100	100	100
BLAGDON			Percentage Difference	0.61	0.52	0.6	0.62	1.4	0.25	0.43	0.08	0.39	0.02	0.36	0.12	0.61	0.46	0.17	0.5	0.22	0.16	0.55	0.25	0.13	0.01	0.22	0.05	0.49	0.14	0.55	0.39	0.59
BODMIN			Percentage Difference	73.49	73.37	72.56	73.34	73.5	73.39	73.35	73.46	73.41	73.37	73.55	73.47	73.21	73.31	72.7	72.36	72.27	73.42	78.51	81.34	73.69	76.89	73.72	73.45	73.6	73.5	73.64	73.91	73.45
BOVEY TRACEY			Percentage Difference	1.59	0.57	1.6	1.38	1.92	1.56	1.54	1.54	1.51	1.97	1.03	0.6	0.41	0.5	0.59	0.71	0.47	0.12	0.8	0.14	0.12	0.34	0.49	0.51	0	0.04	0.13	0.21	0.1
BOVER ASHTON			Percentage Difference	2.73	3.09	2.13	1.23	1.22	1.24	0.96	1.63	2.02	2.38	4.05	0.69	1.51	0.53	1.14	1.51	1.33	0.45	0.45	0.69	0	0.24	0.7	0.68	1.09	0.89	1.65	2.34	1.91
BOVHAYS CROSS BSP			Percentage Difference	32.7	33.43	35.12	35.13	37.05	34.3	34.62	32.59	33.37	33.03	31.99	33.15	35.16	35.91	36.3	36.32	36.79	37.41	36.63	35.8	34.9	27.41	24.4	23.02	21.32	20.35	19	27.35	21.45
BRADLEY LANE			Percentage Difference	7.76	7.93	4.08	4.72	4.54	5.04	5.71	6.1	7.39	7.68	7.02	6.82	5.93	4.39	2.99	2.65	2.68	2.51	2.33	2.51	2.51	2.08	2.46	2.44	2.69	2.58	2.82	2.39	3.12
BRADLEY STOKE 132KV			Percentage Difference	25.62	26.05	25.83	25.55	25.4	25.51	25.5	26.31	25.87	25.63	25.42	25.8	24.54	23.44	24.26	23.81	24.16	22.9	22.39	22.38	22.53	23.08	22.72	18.6	19.02	18.7	18.71	18.82	18.39
BRADLEY STOKE			Percentage Difference	3.26	3.16	2.7	4.72	4	3.69	2.59	3.31	3.39	3.65	4.02	3.53	3.45	4.42	4.44	4.64	4.72	4.44	4.23	4.52	4.6	4.2	4.6	4.67	4.22	3.93	4	4.29	4.09
BRATTON FLEMING			Percentage Difference	3.15	1.04	0.93	1.56	0.25	0.67	2.96	1.32	1.41	1.16	1.62	0.52	1.12	0.2	0.93	0.11	0.2	0.34	0.56	0.47	1.47	0.69	3.09	0.25	0.44	0.69	1.04	0.39	0.41
BRAUNTON			Percentage Difference	27.72	27.39	28.28	29.34	30.08	30.28	29.07	30.46	30.06	29.14	29.12	30.23	29.32	28.21	28.55	31.68	34.89	32.76	34.55	42.65	42.52	45.44	43.07	41.68	39.97	44.7	43.22	46.46	48.36
BRIDGE MILLS			Percentage Difference	2.44	1.96	2.14	0.33	2.13	1.71	1.17	2.19	1.51	2.39	2.36	0.3	7.85	2.12	6.89	8.3	9.77	3.15	1.86	10.65	10.57	13.88	13.21	13.17	12.94	14.04	12.13	9.87	1.9
BRIDGWATER GRID BSP			Percentage Difference	5.26	5.19	4.44	4.73	4.68	4.24	4.32	4.34	4.15	3.94	3.86	3.81	4.22	3.92	3.03	2.43	2.81	5.68	5.58	11.58	20.44	30.8	32.58	34.22	13.62	29.52	34.04	44.42	38.71
BRIDGWATER LOCAL			Percentage Difference	0.08	0.72	4.29	2.96	0.23	2	2.74	2.39	2.9	0.05	1.16	2.33	0.56	5.96	4.12	5.65	5.79	6.4	2.21	3.9	4.3	7.39	2.5	3.44	2.8	3.56	3.84	2.48	2.36
BRISTOL INTERNATIONAL AIRPORT			Percentage Difference	10.16	11.12	6.45	6.45	10.16	10.16	8.73	6.3	7.45	6.79	6.36	5.25	6.96	4.88	5.77	5.34	6.6	5.26	1.28	2.94	1.84	3.22	2.09	2.55	1.02	0.89	0.71	0.39	0.08
BROADWEIR			Percentage Difference	3.66	4.49	2.02	4.78	6.5	4.2	2.29	6.77	4.75	3.5	4.15	5.03	2.96	1.08	1.04	1.54	0.74	1.33	1.16	1.36	0.12	0.37	0.19	0.39	1.4	1.21	1.09	1.07	0.91
BRWA 132 kV			Percentage Difference	45.02	43.58	54.3	57.84	123.8	132.8	145.8	146.4	137.5	132.9	157.1	156.5	91.21	172.2	118	82.52	34.85	40.09	33.89	27.13	67.98	106.2	122.6	101.4	133.7	109.8	76.23	52.15	43.79

Heat Mapping Data Representation

A number of examples of classic heatmaps are shown below[3]. The layout of these are all formatted as being one year extent (down the plot each line being derived from a single PowerOn dayfile) by time of day (across the plot, based on the 48 half hour average data points contained within each PowerOn dayfile). A number of examples of different types have been illustrated to show the different forms of the plots, for example: the wind farms do not (as would be expected) have a diurnal dependency whereas the solar farms have a very clear diurnal profile. The client profile shows very clear inactive periods (holiday shutdowns at Easter and in August) and highlights weekends. On the voltage stability plot, there is an indication of a diurnal and seasonal variation, while this is very clear on the BSP and primary transformer plots.

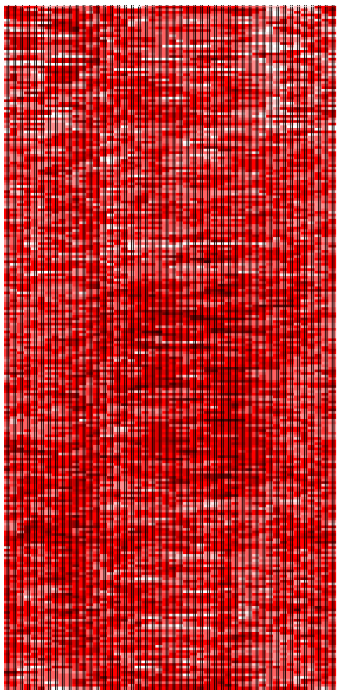

Figure 14 - Heatmap - Voltage Stability at Abbey Wood Primary

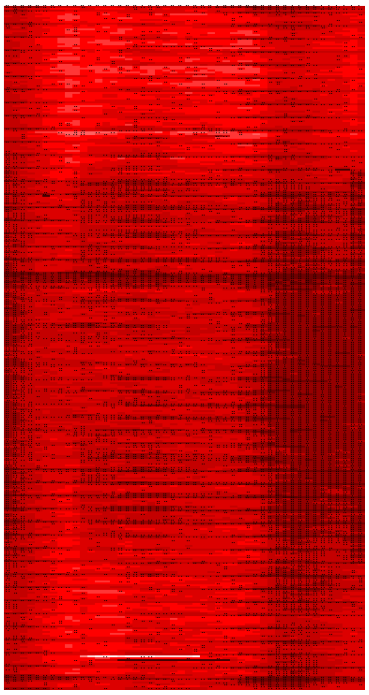

Figure 15 – Heatmap - Foxhills Primary Transformer

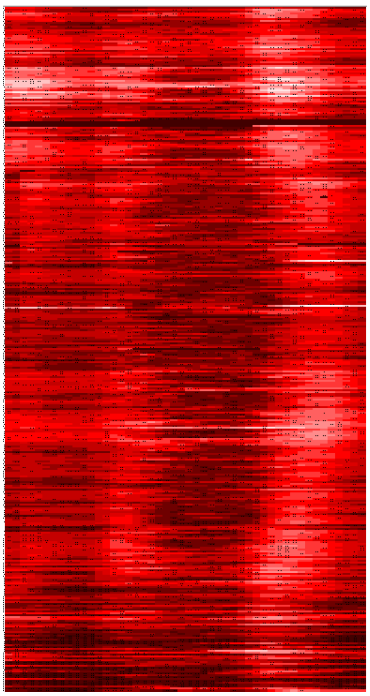

Figure 16 – Heatmap – Fraddon BSP

[3] For further more in-depth information and analysis see Ref [7].

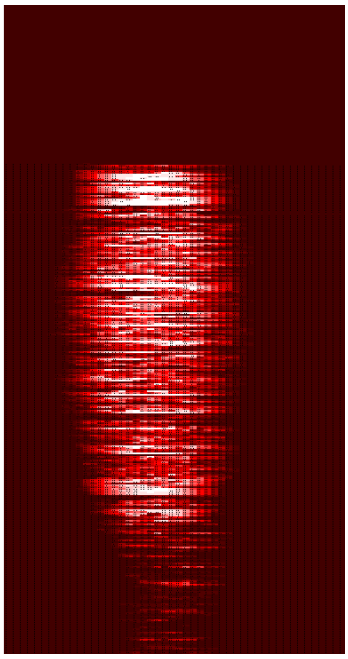

Figure 17 - Heatmap – 33 KV Solar Park

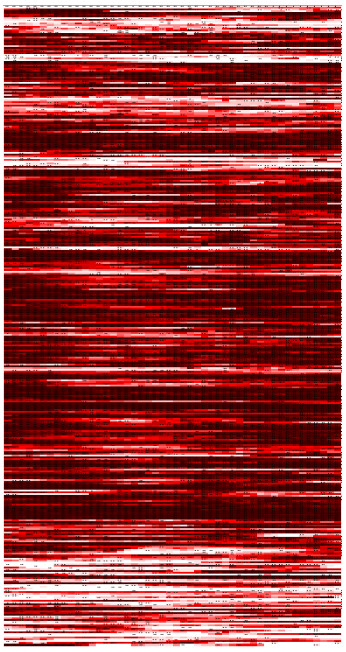

Figure 18 - Heatmap – Wind Farm

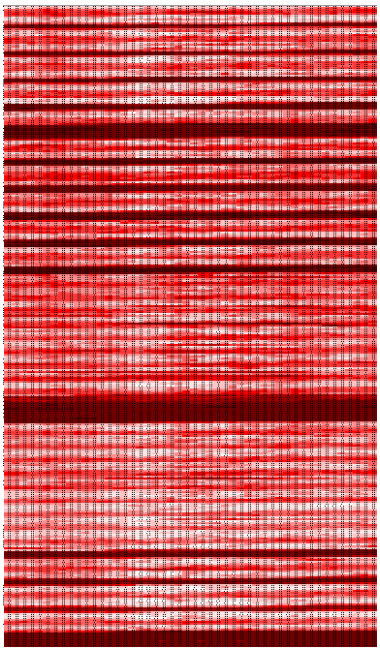

Figure 19 – Heatmap – HV Client

Cross Reference Analysis Results

This table shows the results of the cross referencing of the PowerOn and Datalogger file sets, matching individual analogues and looking for matches, match failures and duplicate entries.

Region	PowerON Entries no Datalogger[4]	PowerON Entries with Datalogger duplicates	Datalogger X-Ref rows	PowerON rows at 31/12/2015	Datalogger Entries no PowerON[5]	PowerON Data all Zeros	PowerON Data all Integers	PowerON Data Some Integers	PowerON Stuck Values
SW	95	55	9603	8737	937	1206	3482	872	23
SW (%)	1.08%	0.63%	109.9%	N/A	9.75%	13.8%	39.9%	9.98%	0.26%
Midlands	7768	14	16074	21458	2554	5332	6216	1764	252
Midlands (%)	36.2%	0.0007%	74.9%	N/A	15.9%	24.8%	28.9%	8.2%	1.1%
Wales	620	185	5390	5148	854	772	1382	624	34
Wales (%)	12%	3.6%	104.7%	N/A	15.8%	14.9%	26.8%	12.1%	0.66%
Total	8483	254	31067	35343	4345	7310	11080	3260	309
Total (Percent)	24%	0.007%	87.9%	N/A	13.99%	20.7%	31.3%	9.22%	0.87%

[4] The Datalogger elements list against which this check is performed is the *PowerON Cross Reference Table*. Some entries cannot be held in this table, for example the Datalogger Derived Parameters.

[5] Percentage of all Datalogger entries. All other percentages are relative to the total number of PowerON entries.