WESTERN POWER DISTRIBUTION
INNOVATION

NEXT GENERATION NETWORKS

Network Assessment Tool: Interim Development Report

Electric Nation

Report Title	:	Network Assessment Tool: April 2018 Progress Update
Report Status	:	FINAL
Project Ref	:	WPD NIA 013 Electric Nation
Date	:	10.05.2018

Document Control

	Name	Date
Prepared by:	Laurence Elner	25.04.2018
Reviewed by:	Nick Storer	10.05.2018
Approved (WPD):	Mark Dale	10.05.2018

Revision History

Date	Issue	Status
25.04.2018	0.1	For EA Technology Review
25.04.2018	0.2	For External Review
02.05.18	0.3	Review comments
10.05.2018	0.4	Final

Table of Contents

Glossary

Abbreviation	Term
Adjacency Model	Pathfinding, redistricting, allocation
BaU	Business as Usual
C#	.NET framework based object-oriented coding language
Coincidence Model	Topological overlay, intersection analysis
Convex Hull	Geometrical spatial analysis method
ERD	Entity Relationship Diagram
EV	Electric Vehicle
Feeder	A circuit which feeds electrical energy from a substation
FMEA	Failure Mode Effects Analysis
Geometric Model	Distances between points, buffers and perimeters
MPAN	Meter Point Administration Number
NAT	Network Assessment Tool
NOP	Normally Open Point
NCP	Normally Closed Point
Raster Data Model	Matrix of pixels (i.e. image based)
REC	Regional Electricity Board
SQL	Structured Query Language
SSIS	SQL Server Integration Services
UI	User Interface
Vector Data Model	Data stored as co-ordinates
WMS	Web Mapping Server
WPD	Western Power Distribution

1 Executive Summary

The ongoing development of the prototype Network Assessment Tool (working title) described in this report is being developed by EA Technology as part of the Electric Nation project. This report describes the progress and developments to date from the previous report in January 2018.

EA Technology has continued the development activities as per the previously stipulated developmental timelines to complete the bulk data upload and processing. Substantial work has been undertaken to set up automation procedures for validation and error control, as well as speed optimisations to allow for the successful processing of the entire dataset for all four licence areas of WPD.

In parallel with this development, further reviews of extended samples and the estimation processes have taken place to ratify results against wider data sample areas to capture variations across licence areas. Further issues were found which weren't observed in the original sample and additional refinements were put in place to overcome these data issues, where possible.

Following testing with sample data from each of WPD's license areas, bulk data (all available data across WPD's license areas) upload and processing was developed, utilising the sample data upload/processing methods, and refined. At this time, preliminary results are that 73% of all substations across WPDs four license areas have been successfully translated (networks mapped), of which 99% were successfully processed by DEBUT (i.e. DEBUT did not report a data input error, the DEBUT assessment results will include failures where, for example, load analysis cannot be undertaken owing to "no customers", "orphan cable segment", etc).

In addition, design work has been furthered for the EV analysis at a targeted level and this is now ready for implementation. Once implemented the outcomes will enable the next steps of design work for wider-scale EV analysis and for the solution module. Additional user interface developments are also required to facilitate the operation of the EV analysis as well as the reviewing of the results by the user.

2 Introduction

This report details the ongoing development of the Network Assessment Tool (working title) since the last progress report (January 2018). The tool aims to provide LV network planners with a new platform to view and assess LV networks under future Electric Vehicle (EV) market scenarios and assess the potential benefit of using smart charging as a method to delay or avoid the need to reinforce networks at risk from EV charging loads.

The tool is currently under development; development has been phased into three distinct workstreams; data transformation and pre-processing; the user interface and the calculation engine. A full development update on the progress to date is provided within this report alongside a status update for each development path.

2.1 The Electric Nation Project

Electric Nation is the customer-facing brand of CarConnect, a Western Power Distribution (WPD) and Network Innovation Allowance (NIA) funded project. WPD's collaboration partners in the project are EA Technology (the authors of this report), DriveElectric, Lucy Electric GridKey and TRL.

Electric Nation, the world's largest domestic electric vehicle (EV) trial, is revolutionising domestic plug-in vehicle charging. By engaging 500-700 plug-in vehicle drivers in trials, the project is answering the challenge that when local electricity networks have 40% - 70% of households with electric vehicles, it is estimated that at least 32% of these networks across Britain will require intervention.

A parallel activity as part of the project is the development of a Network Assessment Tool, this aims to enable a LV planner to assess smart charge solutions to support plug-in vehicle uptake on local electricity networks. A key outcome will be an analysis specifically tailored for highlighting plug-in vehicle related stress issues on networks and identifies the best economic solution where appropriate. This 'sliding scale' of interventions will range from doing nothing to smart demand control, from taking energy from vehicles and putting it back into the grid, to traditional reinforcement of the local electricity network where there is no viable smart solution.

The immediate challenge to such a tool is the prevalence of poor data quality historically present for LV networks in comparison to the vast and accessible datasets available at HV levels. As such, the tool under development will be of great interest country-wide as the next step to high visibility of LV network data at the planning stages. The outcomes of this project will be communicated to central government and the GB energy and utility communities.

This report focuses on the developments undertaken since the previous reporting cycle. Namely this is focussing on the data sample algorithm testing and full processing to import and calculate all four of WPDs licence areas.

3 Summary of Previous Progress

In the span covered by the previous report (August 2017– January 2018), development of the NAT had continued to implement the initial user interface and improve processing and reconstitution of the network data ("translation") such that load flow analyses could be run on successful networks. Common failure modes were manually analysed and a mechanism to automatically flag these successful networks was then designed.

The next development stages identified in the previous report were:

- To establish a method for scenario-based EV analysis across local networks, and the user interface requirements this would require.
- To review and progress upon the estimated network model for failed analysis attempts.
- To implement the FMEA approach as described in the report.

The following sections will detail the tasks performed in widening the sample data processing and assessment on to full scale data processing and assessment, following onto to other advances made and developments undertaken since the previous report.

4 Overview of the Latest Progress

The primary developmental focus on the Network Assessment Tool (the tool) through to April 2018 has been on realising the plans described in the previous report. Since January 2018, development work has primarily been focussed, in line with the development plan timeframes, to have a full bulk upload completed, that is uploading, translating and assessing available data from WPD covering all four of WPD's license areas.

Extensive testing iterations have been completed to further ratify extended samples within each licence area to attempt to uncover any new issues which may have not been present in the initial sample in Plymouth alone. As such, initial executions of extended data samples, one sample area per licence area, were completed before attempting to load the entire dataset sample for all four licence areas. This approach allowed an extended failure mode analysis to be run in parallel with the development work required to upscale and complete a full processing batch process and additional processing speed optimisations.

As anticipated, additional regional issues were found, upon which further work has reviewed and refined the algorithmic spatial procedures for accuracy optimisation across the wide-ranging network topologies and various data issues which can occur due to missing or incorrect and/or outdated data.

4.1 Initial Spatial Processing Refinements

There have been several improvements to the algorithm along with the updated process flow, mostly within the Spatial Processing section, in order to get the highest success rate of networks processed.

The most significant of these has been the change from processing a substation while taking into consideration cable assignments to surrounding networks, to completely independent substation network processing. This prevents errors or certain assignments in one network translation cascading over and breaking what would otherwise be a successful network translation on a neighbouring substation. Such events were highlighted in the previous report as a NXTDR error, often caused by islanded cables, further-than-ideal NOP placement, and other such problems. Processing networks independently resulted in a much higher success rate and often better-quality networks, as each site can be processed considering only features relevant to that site.

Occasionally, this approach does result in a cable segment(s) being assigned to two separate networks, that would not take place in reality, but this does not affect the DEBUT results by any noticeable degree. This is usually because one of the assignments of the cable (to a particular substation network) has no customers attached to or beyond it, and the cable has no current flowing in it; it is an "extraneous" cable and can be safely ignored.

Another addition to the algorithm has been a mechanism to prevent loops within a feeder, which was a major cause of errors at the time of the previous report. Most loops are naturally occurring within a feeder's cables, presumably with a NOP included in reality. A few are accidentally caused by the algorithm in complicated networks. Either way, DEBUT is not able to process such networks and an error is generated. To prevent this, the mechanism traverses each feeder outwards from its start segment, making a note of which cables have been traversed. Upon traversing a cable for a second time, i.e. a loop in the network, the mechanism then de-assigns this cable from the feeder.

The new processing format has resulted in several of the failure mode categories detailed in the previous report to be rendered obsolete. In particular, these are the NXTDR and NOP categories; failure caused by error in processing a neighbouring network, and incorrect inter-network NOP assignment. As detailed above, protection has also been added to prevent loops in the network, rendering the two interconnection-based error categories also unnecessary. These changes resulted in many previous network translation errors becoming successful; the remainder being re-categorised under a different error type. In addition to this, the categories OUTS and SUSP (problems caused by outliers and incorrect DEBUT results) have also been folded into other error categories as there was no benefit in categorising these separately, in particular as the SUSP category has been resolved by mapping across the correct transformer sizes.

4.1.1 Final Success Rates in Plymouth Before Sample Expansion

With the improvements outlined in section 4.1 above, a stark increase in the overall success rate was noticeable. The success rate (proportion of GOOD sites) across the Plymouth sample in the previous report was 32%. Table 1 below details the revised results for the Plymouth sample area. To note, these were all manually assessed to the same failure modes and methodology as previous reviews to enable an accurate cross comparison of the algorithm iteration.

Category	Issue	Sample Volume
NFEEDS	Problem due to Number of Feeder Start connections (duplicate, too few or too many)	3.7%
BAD	Unidentified issues and/or miscellaneous	1.3%
BAD-DISJOINT	Disjointed feeders, no route to source substation	4.6%
BAD-COLOUR	Non-ideal feeder segment assignments	9.1%
BAD-ISLAND	Islanded cable	0.5%
BAD-ZERO	No cables assigned to colour group	0.0%
NOCUS	No connected customers to sub	10.0%
FILE	Problems with file generation (i.e. no cables near sub)	11.9%
GOOD	No problem	58.9%

Table 1 - Plymouth Spatial Processing Success Rates

The algorithm is now at the stage where almost 60% of tested substations return full feeder translations, which would match a manual assessment and return reliable results for the network's capacity. A further 5% of sites currently classified under 'BAD-COLOUR' are a result of two feeders in similar spatial positions having their colours swapped but would return capacity estimates very similar to what they should ideally and could be classified as having appropriate results. In addition to this, there are 21.9% of sites which are un-processable due to a lack of either customer or cable data.

Discussions with WPD have realised that these sites with missing data are likely in most cases due to being present/future developments in which data is currently incomplete, or the networks are third party owned and operated. As such these are of little concern to the development process. So, if one were to only consider the networks which DEBUT has the possibility of processing, the algorithm has an 89.7% success rate. EA Technology are planning a "Development" release of the NAT to WPD, by July 2018, that will enable further assessment of such failures, with a view to identifying solutions to such data quality issues.

Further to this, all results provided thus far are at a distribution substation level. With each substation serving approximately 4 feeders on average, some substations which are categorised into a non-success (i.e. not in the 'GOOD' category) may have some feeders which are processed and calculated appropriately. Therefore, at a feeder level, relevant success rates where data is complete are somewhat upwards of 90%.

For these last few networks which the algorithm failed to process, signified by a DEBUT error message, there will be the alternative centroid-based method for estimating the network, as detailed later in this report. Once this alternative is fully developed, an estimate view of capacity will be available for every LV feeder identified.

The change of biggest note in the Plymouth results was the increased success of the algorithm in urban areas. With their dense network population, urban areas were far more prone to the “chaining of errors”, where the results of one network may negatively affect others around it. When sites were processed independently, the success rate within the urban section of the test area was 60.9%. Conversely, the success rate in rural areas was slightly lower at 56.0%: this is due to the higher likelihood of rural substations having either no customers assigned or no cables near them. Having a high success rate in urban areas, as opposed to half that of rural areas as was true in the last report, is highly beneficial; it is far more likely that electric vehicles will be taken up in urban areas than rural ones for the foreseeable future. But, it is recognised that the effect of EV charging is likely to be greater in rural areas as the networks are traditionally weaker. As stated earlier EA Technology are planning a “Development” release of the NAT to WPD, by July 2018, that will enable further assessment of such failures, with a view to identifying solutions to such data quality issues

4.2 Expanded Sample Area Dataset

The initial data sample provided, which covered the Plymouth area, was deemed, by WPD, to be of a very high data quality, in general. This provided a solid testbed of data to develop the spatial processing algorithms. Now that the algorithm was achieving very reasonable results on the Plymouth data, it was decided that the samples used to assess the success of the algorithm would be expanded to four areas, one from each license area. As well as the original Plymouth sample (South West), three additional areas of Cardiff (South Wales), Lincoln (East Midlands) and Worcester (West Midlands) were chosen. The new sample areas were chosen to span both domestic and I&C areas, with some rural outskirts, as to test the algorithm on a variety of different, but likely, situations.

There were several problems immediately obvious on examining the new test sites. Perhaps the most significant of these arose from the reliance on the 'DumbUgCable' database, an historically older data set than others provided by WPD. Which from initial analysis at the beginning of the project was of the third order of reliability due to the much lower quality of populated data fields. Initially this was thought to be cable types only, however, spatial reviews demonstrate the geospatial records are also misaligned in some cases and give a much more scattered appearance overall when directly compared against the Plymouth sample. The ramifications of this are clear, joining cable sections together to form a feeder is harder, the results in these areas which rely on this dataset alone will not be as accurate or successful in network translation and thus network capacity estimation.

Approximately 28.8% of all cables are sourced from this database in the full collated dataset. Figure 1 illustrates the spread of these across the four license areas. The biggest problems will be in the south and west of South Wales, and in the east of the South West region.

Figure 1- Population density of cables from the problematic 'DumbUgCable' database

4.2.1 Success Rates Post Data Sample Expansion

After an initial review of the newly processed results, some immediate issues were identified, where the spatial algorithms weren't performing as well as expected – owing to the more varied quality of data encountered in the new sample areas. Namely, customer location inaccuracies causing outliers, start segments not being correctly identified and spatial gaps between segments were found which weren't present in the Plymouth area. To overcome these issues additional modifications were made to the network translation algorithm and the spatial algorithm stack was iterated.

The following refinements were put into place;

- Customer spatial outlier protection re-introduced.
- A refinement to the previous customer outlier protection to establish a suitable boundary distance.
- Start segment protection changed from closest N feeders to considering all potential feeders in a 25m radius and picking the most appropriate via the convex-hull intersection method as previously covered.
- 0.5m tolerance between two cable end connections successfully implemented for traverse of networks during feeder mapping stages. Noting that this modification increases execution times and will only be executed as a last resort step for data translation before going to full estimation.

The success rates for each area are given in Table 2 below;

Category	Issue	Sample Volumes (%)			
		Cardiff	Lincoln	Plymouth	Worcester
SHORT	Disruption due to non-connected segments	1.4%	1.4%	0.0%	0.5%
NFEEDS	Issue due to the number of start segments	14.2%	4.6%	2.3%	2.3%
BAD	Other cause/unidentifiable problems	0.7%	0.0%	0.4%	0.6%
BAD-DISJOINT	Disjointed feeders, no route to sub	0.0%	0.0%	1.8%	0.9%
BAD-COLOUR	Strange/non-ideal assignments	7.8%	10.5%	11.0%	6.8%
BAD-ZERO	No cables assigned to a group	0.0%	0.5%	0.0%	0.0%
NOCUS	No connected customers to sub	0.0%	11.9%	10.0%	12.3%
FILE	File generation error (no cables)	4.6%	4.1%	13.7%	12.3%
DATA	Data issues, e.g. consumer type errors, zero length segments	1.4%	2.7%	0.0%	1.8%
LINKBOX	Short feeders/misidentification caused by link boxes	26.5%	20.5%	0.0%	10.0%
BAD-DATA	Extremely bad data i.e. no route to customers, too few starts, inaccurate spatially	8.7%	0.9%	1.4%	2.3%
GOOD	Success	34.7%	42.9%	59.4%	50.2%

Table 2 - Expanded Data Sample Results

Largely the outcomes were as anticipated, lower in areas where the data quality is of a lower standard when compared against Plymouth. It’s reassuring to note, that other areas where the quality is of a good standard, the algorithm is still performing very well as Worcester demonstrates. Cardiff was the lowest performer; however, it should be noted that a failure of network translation is the ruler, some networks present in this area upon a manual review are untranslatable even by the human eye. The volume of these are, however, unquantifiable.

A secondary outcome of this work is that areas such as Cardiff and Lincoln to some extent, will be highlighted as being areas of poor data quality. WPD will subsequently be better positioned to prioritise their Data Improvement programmes to target these areas first.

After the data samples were extended and reviewed to a satisfactory level, it was decided to continue with the full data load and processing of all four of WPDs licence areas.

4.3 Bulk Data Upload

4.3.1 Overview of the Enabling Developments

EA Technology’s NAT development work has recently been focussed on the batch bulk data uploads and processing for all four licence areas of WPD. Previously established was an upload routine which was essentially a one shot small upload process designed to handle sample sized volumes only. This also required some manual setup work to configure before execution. Therefore, the developmental works required for a bulk data upload were as follows;

- Initial extensions to cope with extended data sample areas which include new data validation configurations for the wider area data variations.
- An initial re-package of the latest algorithm iteration, version 0.11 for integration into the batch process. Note, there are two parts, a primary and secondary routine. At present only, the primary routine is being incorporated for review. This is due to internal thinking that this process may be redundant, it also has heavy resource implications.
- Upon an initial test this was found to be performing in substandard processing times, considering the volume of data which was to be processed. As such, there was a follow-on concern that upon attempting a mass data load an unexpected failure at any point (such as new requirements for validation checks which are frequent in loading new datasets for new areas) would be irrecoverable and too much time would be lost. Thus, to avoid loosing 1-5 days of processing time, new optimisations were considered and implemented in order to cope;
    - Speed optimisations; namely breaking up of the algorithms into smaller segments which could be called as a service for a given substation or group of substations. This allows not only for a failure to take place, but for the specific substation/group of substations to be flagged as an issue occurrence. A secondary benefit is the ability to parallel process these blocks of code to further speed up processing times.
    - Ancillary code and methods to iterate through and break up the whole dataset into manageable batches.

Re-execution of the whole dataset has been completed against the latest algorithm version 0.11 (primary routines only).

The executions of the DEBUT calculations follow suit and are similarly handled. This has just been completed at time of writing this report, further work is required to assess the results.

4.3.2 High-level Results of Mass Data Processing and Calculation

The level of granularity of the available success and failure metrics as discussed in the previous sections are not available when processing the wider-area data set. Fundamentally this is due to the previous being the output of manual reviews and the process is now automated for processing the bulk data. This isn't expected to improve until the confidence metric is implemented and tested.

At the time of writing, the metrics available are direct outputs from the bulk data processing, which are;

- Failure of initial spatial validations: when the data is ported from the staging database into the spatial area. This filters out any glaring issues, such as customer data with no locations, substations which are not contained in our assets data or miscellaneous invalid entries which can cause issue with this routine.
- Spatial processing/DEBUT input file generation: network translation has failed to generate a valid input file for DEBUT to calculate. Several issues cause this, primarily these are invalid network data points which the algorithm can't cope with.
- DEBUT calculation error: direct calculation errors with associated error messages. A varied mix of errors which can be further split out, however due to the volumes of these only being 743 substations, these are negligible at present. These are similar causes to the previous spatial processing errors, however in a more minor form as the input file was successfully generated.

Figure 2 below, illustrates the success rates that have come out of the initial bulk data processing.

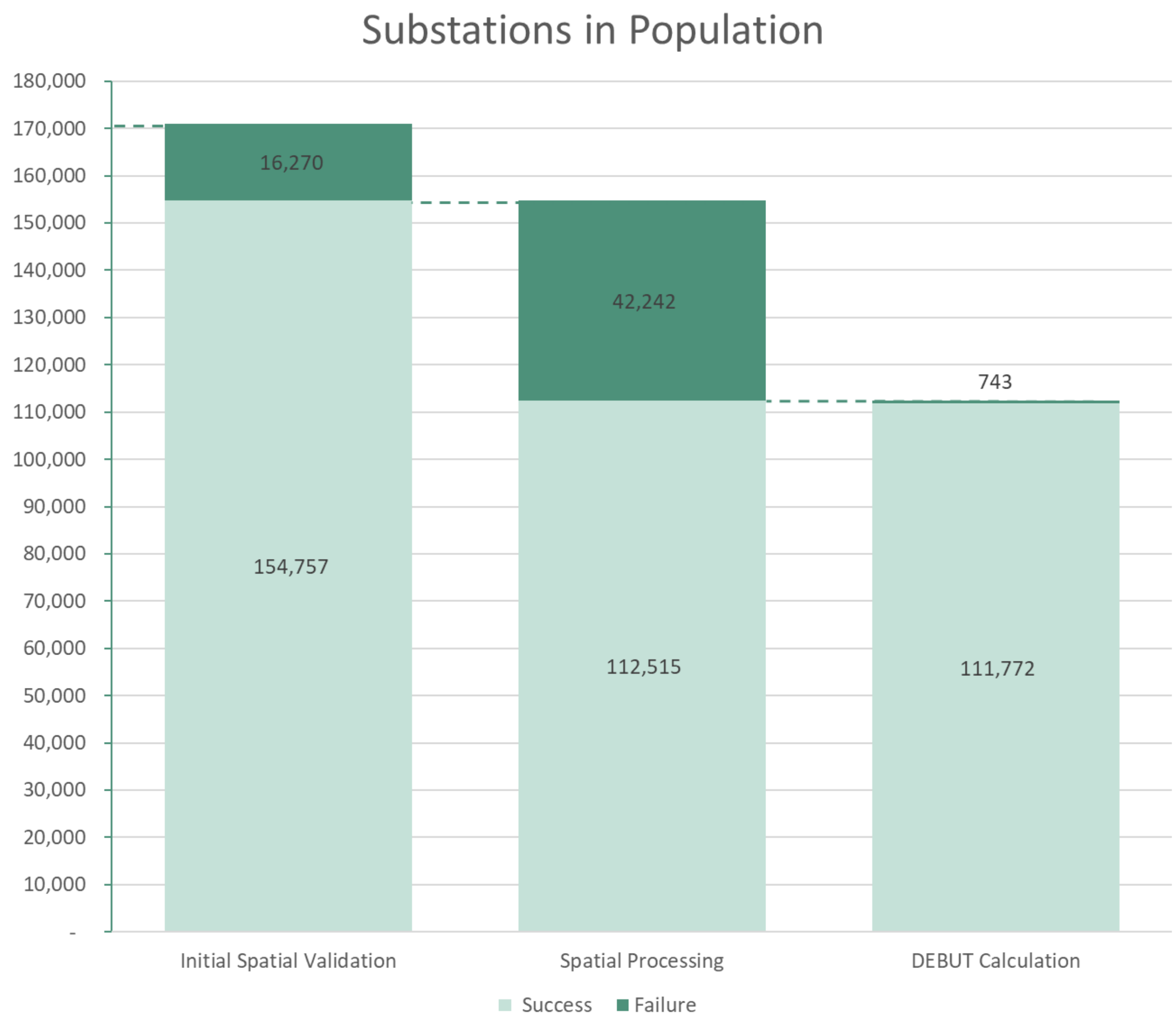

Figure 2 - Bulk Data Processing Success

In summary, that's a very encouraging 65% of all substation IDs listed in the customer database as having a successful translation and calculation. It should also be highlighted, that 9.5% of the failed attempts from the initial validation are un-processable due to absent/invalid spatial data being available for the customers. In addition, a reasonable portion of substations within the secondary stage of failures are also attributed to networks with missing data which are due to incomplete data due to third party owned networks and future developments which brings the accuracy up even higher, estimated to be around 75-80%.

4.4 Development of a confidence metric for individual networks

The sharp increase in the success rate after the conclusion of algorithm iteration 0.7 prompted a re-think in how the FMEA would be implemented. There were 8 automatable categories in the original design: two were since eliminated and a third reduced so far in volume it had become negligible and barely worth the processing to isolate it.

It was decided, that what would be more useful to the end implementation of the NAT would be an automated measure of confidence in the resulting network, as opposed to a discrete category of failure, especially as the five remaining categories were quite general. The aim of this metric is to assess how many underlying assumptions have been made in forming the feeder setup of the network and assign a value to the network based on the types of assumption used. The more assumptions are used, the more likely it might be that its cables have a poor assignment and the less confidence there is in the resulting capacities of the network.

A network which has obviously failed in translation (i.e. results in an error in DEBUT processing) will progress to using the network estimation method and automatically be given a low confidence rating. Networks which have less obviously failed (e.g. success in DEBUT but has swapped feeder identifications for some stretches of segments) will have various features and assumptions highlighted by the confidence process, as such the confidence rating will be lowered.

There are various assumptions which can be used throughout the process. The list so far includes:

- Use of the network estimation method as opposed to real cable data (i.e. DEBUT fail).
- If customers assigned to the substation have assumed locations.
- Presence of NOPs between two feeders on the same network.
- Use of the interconnection protection algorithm.
- Presence of outlier customers in a group, >3σ (standard deviations) away from the group's centre.
- How much two customer groups overlap with each other, which has been identified as a primary contributor to misidentification of segments to feeders ('BAD-COLOUR').
- If customers are particularly far away from their feeder cables.
- Quality of the source data; original data source table and any missing attributes.
- Whether a spatial tolerance is required to connect cables of a feeder to connect to each other (e.g. where cable end/start coordinates to not exactly match, but are within (say) 1m of each other it could be assumed they are connected).
- The customer/segment ratio, a measure of how spread out the customers are.
- Presence of industrial customers, and the accuracy of the demand profile.
- If any link boxes are present, as there are no fixed rules for how these are connected.
- No customers or cable data available.

Each of the factors to consider will be given a weighting. If a network incorporates a certain assumption when reconstituting the cable data, then it has the corresponding number of points deducted from its confidence rating, until a minimum of 0 points is reached. A network with no assumptions will have a maximum confidence rating, which is currently 100. The exact weightings to be allocated to each assumption are yet to be determined.

4.5 Assessment of the Centroid Based Network Approximation

The Centroid-based Network Approximator (CNA) algorithm, described in the previous report, is a way of obtaining an estimate of the capacity of a network in the event that the actual cable data fails to have been processed successfully. This could be due to a lack of available cable data, or a case where the main algorithm has failed to produce a valid DEBUT network. Referring to the previous section, this will be used to process the remaining 25% of networks which failed to process and translate into a valid DEBUT network.

The algorithm works by creating, for each feeder's group of customers, a main branch line to the centre of the group's convex hull, and subsequent branches in different directions out from this centre until all customers have been assigned to a branch.

4.5.1 Estimated Manual Proof of Concept

The CNA was in a state of manual assessment at the time of the previous report. This involved taking networks successfully processed by the main algorithm and by DEBUT and comparing the successful network to the scenario where the CNA was applied to this network instead. This process was undertaken in WinDebut to model both the successful and a manually interpreted CNA network. The use of known results acted as a control group to assess the accuracy of the estimated method.

The manual analysis used a few approximations to speed up the analysis:

- All customers were assumed to be of the demand profile class 'Unrestricted Medium Consumer (URMC)' type
- All cables were assumed to be the same type, which was determined by taking the modal cable material from the cables in the original network, with the size of all cables being the largest used in said network
- Cable lengths were to 1m precision, limited by the measurement tool on the UI
- Customers were assumed connected to the end of an estimated cable (end node) if they were located within a 30m locus of the length of the cable.

The success of the CNA was determined by comparing some calculated capacities (maximum voltage drop, average 1st leg cable capacity, and maximum transformer demand) to those from the original Debut run. 16 networks underwent this manual analysis, all from the Plymouth sample area and which were deemed to have successful original assignments by manual analysis.

Voltage drops were the least accurate assessed capacity, being out on average by -20% with a standard deviation of 26% (percentage change, as opposed to percentage voltage drop). This could be due to a variety of reasons, such as difference in cable lengths, where customers were assigned within networks, as well as the difference in customer types and cable types. However, due to the time needed to manually build CNA networks in WinDebut, not enough sites were processed to draw a significant conclusion on which factors were the biggest contributors, or if any of these could reliably be used to adjust the estimated voltage drop to a more accurate number at this stage.

The average cable capacity in the 1st legs of the feeders was on average out by -8% with a standard deviation of 16%. This was most likely due to the cable type assumption. Originally, the automatic cable assignment feature in WinDebut was used to assume a type for all cables, but this was often so different to the cable types in the real network that the capacities became completely different. Hence the cable type assumption described earlier was developed; this still contains some level of inaccuracy, especially in networks containing many types of cables. There was a slight correlation observable between CNA accuracy (by cable capacity) and percentage of cable in the original network being of the assumed material. If proved true on a larger scale, this could be incorporated into the confidence metric in some form.

Maximum transformer demand was, on average, out by only 1%. This was directly correlated to the difference in percentage of customers that were URMC type in reality. The approximated network at this stage assumed all customers were URMC type; the discrepancy in maximum demand was almost certainly due to this assumption as customers of different types would have different demand profiles to URMC. Mostly these different customers were Economy 7 profiles, which have smaller maximum demands than URMC and hence explain the slight demand overestimate.

It should be noted that this method is only a transitional aid while WPD complete their parallel works to improve and refine their datasets. With further refinement, calibration and the addition of some logical compensation factors, it is hoped that the method will enable previously un-processable sites to have an estimate benchmark which can be further studied and compared.

4.5.2 Implementation Results

The method of the CNA was coded as per the previous framework, as opposed to including any of the assumptions used in the manual test. For validation purposes, the CNA was applied to every network after the original DEBUT analysis. The original capacities of all networks manually determined to be ‘good’ implementations were compared with their CNA counterparts. In total, 76 networks were analysed in this way.

Initial reviews showed disappointing results due to an implementation issue whereby the original manual analysis used a 30m locus of the cable, however, the coded implementation used only a 30m radius of the end node. This causes the network to often switch back on itself, caused by the presence of customers along the length of the cable but not at the node itself, shown in Figure 3. From reviews, this is the key difference which has caused the discrepancy in accuracy between the manual and coded versions of the CNA, as the coded version is sometimes prone to generate far longer cables than it ought to, and reliability becomes more dependent on the spatial distribution of the customer groups.

Figure 3 - An example of the "switching back" issue observed

The biggest change in reliability is in the voltage drops for the networks. The difference across the new data set was on average 4%, but with a standard deviation of 50%. This demonstrates a skewed distribution; the median is -13%. For ease, there was a slight change in metric for cable capacity; in the larger test the maximum cable capacities, instead of average 1st leg, were compared. This gave results of average 14% difference with standard deviation 37%. Again, this is a skew distribution, with a median of 0%.

Regarding difference in maximum transformer demand, the larger-scale test confirmed the difference in actual and estimated demands was entirely due to the assumption that all customers were URMC. As expected, using all the actual customer data, had no difference at all between estimated and actual maximum transformer demands.

A second implementation will be iterated and tested when development timelines allow, correcting the radius assignment of customers to a locus to prevent switching back. Pending further considerations, other measures could also be implemented such as incorporating existing 1st leg cables (if available) to increase accuracy.

5 Modelling EV Uptake Scenarios

Once a baseline capacity has been established for all the networks, the effect of electric vehicles on the networks can begin to be explored. The aim of this development pipeline is to periodically add an increasing number of EVs to the network (either pre-set or with an uptake defined by the user) and study the effects on the networks available capacities and constraints.

This will originally be done on a license-area wide scale, to visualise aggregate effects every half-regulatory period, e.g. the approximate number of transformers, or the volume of cable/conductor segments that will need to be replaced after T years. There will also be the option of more targeted analysis; where a user can assess the effects of various settings and solutions on an individual substation in greater time resolution than on a wider area.

5.1 Baseline capacities

Once the network processing is complete, every network will have a measure of its capacity at the current point in time, year 0. Each individual capacity metric; maximum voltage drop, cable capacity and transformer demand are given Red/Amber/Green codes to indicate their baseline status. Other potential variables such as earth loop impedance headroom may also be considered. The thresholds for each individual capacity are shown in Table 3.

Constraint Metric	Green	Amber	Red
Voltage Drop	< 3.5%	3.5% - 4.79%	> 4.79%
Cable Utilisation	< 80%	80% - 100%	≥ 100%
Transformer Utilisation	< 80%	80% - 100%	≥ 100%

Table 3 - Thresholds for the constraint metrics

Substation objects on the UI will be colour-coded by their baseline capacity RAG, as will the areas covered by Energy Supply Areas and license areas when the UI is sufficiently zoomed out. To establish a consolidated constraint health of a whole substation a simple analytical hierarchy will be employed as illustrated in Figure 4 below.

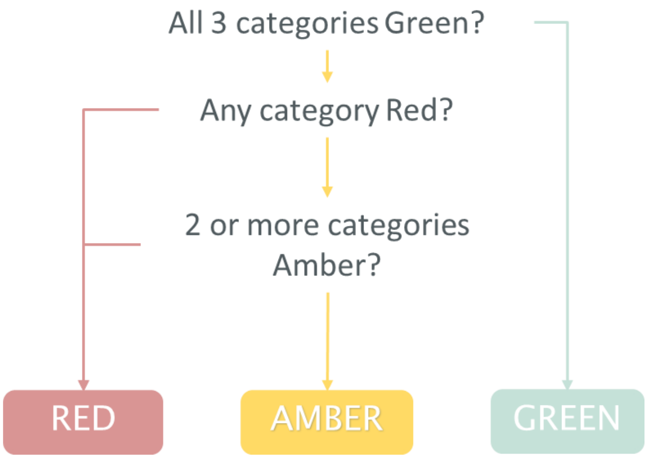

Figure 4 - RAG combined score

5.2 Modelling EV Uptake

The uptake of electric vehicles with time is modelled by a specifically calibrated cumulative frequency graph. There are two inputs: the maximum expected penetration of EVs (Y), and X, the number of years expected to take for this to happen. The expected percentage uptake for each year from the baseline (year 0) to year X is then calculated. The time = 0 penetration of EVs will be the number of EVs currently on the network as according to the available data. An example of this is shown in Figure 5. This is then combined with a known quantity of customers to give an expected number of EVs to be distributed among the customers. The approach can be applied to an individual network, simultaneously to each individual network or substation.

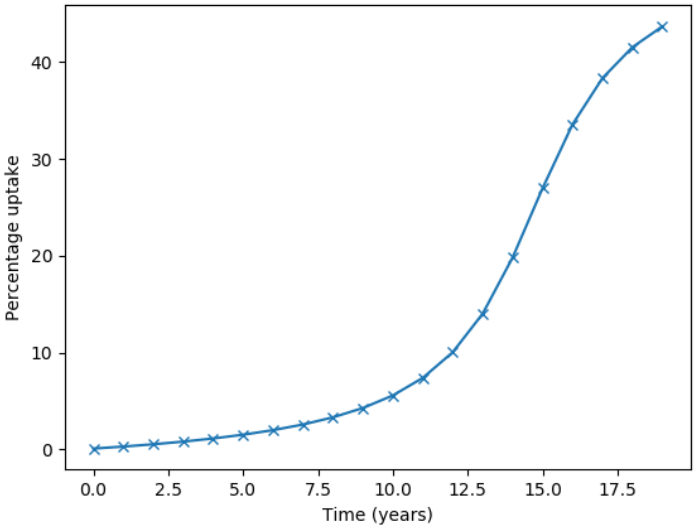

Figure 5 - Example of an EV uptake graph, for a maximum penetration of 50% expected at 30 years, with an initial penetration of 0.1%

The EVs are allocated pseudo-randomly to customers across the network as the years progress, however each deployment will be logic based and the process will be repeatable such that each iteration of modelling the results will not be skewed by specific EV allocations. The default spread of the expected EVs is designed to be approximately even across the network. If the data becomes available, a method to incorporate EV allocation probabilities by socioeconomic factors, such as ruralness, then some degree of natural clustering is to be expected.

5.2.1 EV Demand

Each customer allocated an EV receives an EV demand profile in addition to its normal domestic profile. The shape and magnitude of this profile will be obtained from the Electric Nation charging database and translated into a DEBUT demand profile which will include the diversity levels, also as an output from the Electric Nation analysis.

A key factor to consider which has been incorporated into this design is that the EVs that will be deployed on the networks are unlikely to be identical. The major difference in EVs, from a network modelling perspective, is the charging rate and the battery size. The EV market can be split into four approximate categories by battery size, any combination of which could be added to the model:

- Plug-in hybrid (average 10kWh battery, but can vary significantly, differs from Small BEV by shape of profile, e.g. BMW i8, VW Golf GTE)
- Small BEV (battery size <25kWh, e.g. Mitsubishi i-MIEV, Renault Zoe (2012))
- Medium BEV (25-50kWh battery, e.g. Nissan Leaf, Kia Soul)
- Large BEV (50kWh+ battery size, e.g. Tesla models)

The progressive uptake of each of these categories is expected to be different. For example, plug-in hybrid uptake is expected to be the first fast-rising category, as they are not too different range-wise to petrol or diesel vehicles. Conversely, the initial uptake of large EVs is expected to be very slow at first, due to the expense and limited initial availability of such vehicles, as well as the long time they will take to charge, but it is likely to reach an eventual maximum penetration much higher than that of hybrid vehicles once these problems begin to be solved. It is suggested that an approximate demand profile and uptake curve is developed for each of these categories, and these are simultaneously applied to a network. Figure 6 below illustrates an example of these four uptake curves for the EV typed categories.

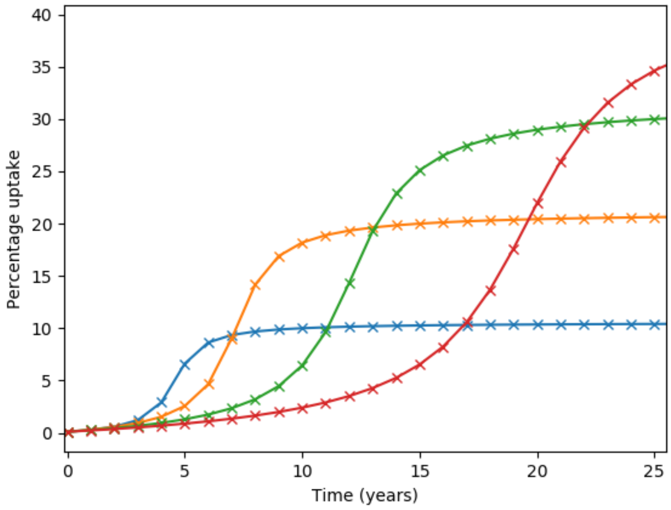

Figure 6 - An example of four uptake curves that could be used simultaneously. Blue: Plug-in hybrids, Orange: Small EVs, Green: Medium EVs, Red: Large EVs

The only limitation to the described method is the execution time required. Not only are the routines required to establish EV deployment volumes and locations, but also divided by type categories, then repeated for each modelled year going forward. This is feasible to be user controlled (manually configured and re-executed directly from the front end) at a targeted substation level, however at a wider-scale, this would take too long to process without substantial dedicated infrastructure to support it.

5.3 Wider-scale EV impacts

Further consideration is still being made regarding calculating this across all four licence areas. Initial thoughts are that the same method is used however the scenario is pre-set and the modelled years are limited to just three as an example. The drawback to this however is that the execution time will be substantial due to the volume of networks and periodic variations. As such iterative reviews become less feasible and the tool becomes less flexible which isn’t desirable.

Optimisations and alternative methods are still being investigated for wider scale reviews. As the baseline capacities are at this point already in place, it may be that with some targeted analysis, the results can be extrapolated to the wider area with some inferred logic and thus allow for a much faster iterations of the wider-scale results.

6 Ongoing Development Path

The foundations of the tool are now in place; the user interface is established and stable, the core data is structured and expanded to all regions (bar, as already mentioned, some apparently missing data), data is spatially processed and all key relationships between the varied data sources have been identified and re-established. Further efforts have gone into refining the optimal algorithm stacks to process, sort and categorise the processing for accuracy and speed. A backstop procedure has been defined and the method to interlace this into the overall architecture has been implemented but is awaiting further refinements for accuracy and calibration.

This section discusses the ongoing developments which are required to provide a functional network assessment tool.

6.1 Remaining Development Paths

The following development paths are still either in progress or awaiting a pre-requisite task;

- Confidence metrics to score each substation network translation based on the available data
- Refinements to the assumption-based estimation of line and load
- Implementation of targeted EV uptake analysis
- User interface additions for EV analysis at a targeted and wider-area results summaries and user interaction are to be designed and implemented
- Methodology for wider scale EV analysis is still to be decided upon
- Options assessment module – which will enable assessment of smart charging as a mitigation method (vs reinforcement) for networks overloaded by forecast EV loads.

6.1.1 Confidence metric

As discussed in section 4.4, the underlying concept has been established. Some final design work to establish the reference values and initial weightings are required before this is ready to be fed into the development timeline.

6.1.2 Refinements to the estimation of line and load for failed network translations

As discussed in section 4.5, the initial implementation is in place, however after a review it was found that there was a misalignment to the original design. A potential fix for this has been decided upon and is awaiting to be implemented for re-execution and review.

6.1.3 Targeted EV analysis and User Interface additions

The next stages for this section are twofold. The method as discussed in section 5 is to be implemented and tested across real networks to verify its fitness for purpose. In parallel with this the user interface will need amendments in order to allow for user interaction and the for the viewing of results summaries, the design for this is under consideration at present.

6.1.4 Wider-scale EV analysis

The design for wider-scale analysis methodologies will be decided upon post implementation of the targeted analysis. This will allow for processing speed implications to be further ratified before committing to further design and development work.

6.1.5 Options assessment

Initial considerations are being undertaken in parallel with targeted network analysis, however further detail is still required to further this design work. In a similar vein to the wider-scale EV analysis, the processing implications are going to influence the feasibility of certain design aspects. As such further work in this area is also pending the initial implementation of the targeted EV analysis.