Smart Network Design Methodologies

Smart LV Design - Functional Specification

December 2019

Contents

0 Document control

0.1 Document history

Version	Status	Issue Date	Authors
1	First Issue	25/09/19	Gruffudd Edwards Charlotte Higgins

0.2 Document review

Name	Responsibility	Date
Francis Shillitoe	Project Manager (NPg)	12/11/19
Alan Creighton	Technical Lead (NPg)	12/11/19

0.3 Document sign-off

Name	**Responsibility**	**Date**
Mark Nicholson	Project Sponsor	02/12/19

1 Introduction

Distribution networks in the United Kingdom are currently experiencing growth in the presence of both low carbon generation and demand, and correspondingly more dynamic but controllable loads – a trend that is set to continue over the coming decades. These changes, along with the improved ability to understand network behaviour through smart meter data, are creating an increasing need to move towards a more sophisticated LV network modelling approach. Indeed, integration of smart meter data as well as an increase in network monitoring at the LV level, and more sophisticated socio-economic spatial modelling of customer load, have the potential to provide significant value to LV network planning and design.

The objective of the *Smart Network Design Methodologies* project is to trial how smart metering data – in combination with other new sources of network-specific data related to demand and generation, and existing network data – can be used to improve the planning and design of the distribution network. This report provides details of the functional specification of a future LV network planning and design tool that utilises the approaches developed in this project's report on Novel Analysis Techniques[1]. We have named the tool "Smart LV Design", but for brevity will be referred to as the 'tool' in this report. The tool would be integrated with Northern Powergrid (NPg) software and IT systems, and would interact with Siemen's secure Energy IP environment, in which smart meter data will be held. In this report we describe functional specifications for a Minimum Viable Product ('MVP'). It would be expected that if the tool were developed the MVP would be tested with a limited set of users and their experiences fed into the development of the second version of the tool. The MVP does not rely on large quantities of smart meter consumption data being available, so development could progress immediately.

1.1 Background

Currently, LV network planning and design (including for new connections) is carried out using several design tools across NPg licence areas as summarised below:

LV Design (DEBUT) (Northeast): LV Design (DEBUT) is an older version of WINDEBUT and is a graphical and analytical tool for LV network study. It performs thermal, voltage drop, fault current and earth fault loop impedance studies on the LV network. These are based on average parametric relationships rather than power flow analysis. For a specific LV network, the user is able to select a specific customer type for each connection, along with either the customer's annual energy consumption – if available – or the average energy consumption for the customer type. The tool translates these into peak demand values for chosen groups of customers on the network, e.g. all of the customers on each feeder, and the network must be designed to meet this demand without violating voltage or thermal limits at any position in the network. In general, the DEBUT method is limited to representing the demand of domestic customers, although NPg has extended the approach to include electric vehicles (EVs) and heat pumps (HPs), but it is not possible to represent generation. The default representation for domestic customers is as the Elexon type 'Domestic Unconstrained', implying that they are subject to a flat price tariff, but other Elexon types such as 'Domestic Economy 7' can be chosen.

Design Demand Calculator (DDC) (Northeast and Yorkshire): This Excel based tool provides the Equivalent ADMD (kW) for any given LV feeder depending on the number of customers and the type of load connected. Apart from normal domestic load, it also takes account of heat pumps and electric vehicle load. Based on the Equivalent ADMD (kW) values, the DDC calculates the total design demand (i.e. total load) in kVA that any supplying transformer or circuit will have to supply. The DDC calculates the Equivalent ADMD (kW) values for typical loads which are assumed to represent the majority of networks, and have also extended the analysis to include EVs and HPs.

[1] Available at: https://www.northernpowergrid.com/asset/0/document/4918.pdf

Network Calc Tool 3 ph (Excel Tool) (Northeast): This Excel based tool primarily is used for motors (starting studies), welders fusing & flicker study. It provides values for the Engineering Recommendation P28 limit, loop impedance, voltage fluctuation, fault current and fuse size.

LV Volt Regulator & Fault Level Calculator (Yorkshire): This is an Excel based in-house tool that calculates the voltage drop, earth loop impedance, fault current and required fuse size.

1.2 Future LV Planning and Design Tool

We have developed conceptual solutions for an innovative future LV planning and design tool ('the tool'), and have demonstrated the methodology on real LV network case studies. The approach is based on advanced statistical techniques that model both customer loads and – via a model of the network and AC load flows – corresponding network states. It incorporates the following features:

- Sophisticated data-driven statistical modelling;
- Risk based analysis – identifies the level of network demand that results in exceeding asset capacities and statutory voltage limits, their impact and their frequency;
- Can be extended to capture the statistical properties of multiple, correlated demands - for situations where the state of a network component cannot be determined with sufficient accuracy by a single aggregated demand;
- Takes particular care to ensure that the high and low extremes observed over multiple years are accurately modelled;
- The ability to carry out analysis in a true power system modelling environment, used economically to manage computational resource requirements;
- Dynamic and flexible, so that it will update to incorporate new data, such as increasing availability of smart meter data, LV network monitoring data and learning about the usage patterns of new technologies and their uptake.

This approach will facilitate more efficient network planning in a changing energy system, maximising use of increasingly granular customer and network monitoring data.

1.3 Purpose

The goal of the future LV planning and design tool is to offer an alternative to the DEBUT and the Yorkshire ADMD spreadsheet tools. The replacement of legacy tools would be justified on account of the new tool's ability to flexibly leverage new data sources, to produce more accurate predictions – due to the use of AC load flow calculations, and to allow managers to define (and change) the precise risk-based metrics they wish to be routinely calculated. These features would facilitate the development of economical, efficient and co-ordinated LV networks.

It is essential that the tool would fully integrate with existing and new NPg IT systems such as Spatial and Siemens Energy IP, i.e. the smart meter secure data storage and extraction system.

2 Overview

2.1 High Level Functionality

The operation and basic functionality of the proposed new tool may be summarised at a high-level as follows:

1. For a specific LV network, the DNO's complete knowledge of the connected customers' existing load is translated to a **probabilistic model of demand**. The tool initially uses what it already 'knows' about demand distributions for a similar group of customers (i.e. the same number, or the same numbers of different types) to provide an initial 'generic' model. Then, all existing data relating to demand on that specific network - mainly smart meter consumption time series and possibly network monitoring time series – are used to update the generic model and **produce a network-specific model**.

2. This **updating** will eventually be achieved by **adopting a** fully **Bayesian statistical modelling approach**. A benefit of the Bayesian approach is that uncertainties in our knowledge of the customers, e.g. how many have electric vehicles or the number and age of occupants in each house, are explicitly modelled, alongside the more fundamental randomness of electrical demand. These uncertainties are, however, 'collapsed' in a mathematically robust way into single values for e.g. the level of total demand at any given point on the network that will be exceeded on average on only one half-hour period every 10 years.

3. The Bayesian approach requires the existence of some data on the relationship between the number of customers being aggregated (to establish the demand at any given point on the network) and the parameters of their probability distributions and – more challengingly – parameters describing the uncertainty in the probability distribution parameters, given the number of customers and possibly customer types, in future versions. It also requires data regarding the precise statistical relationships between different aggregated demand variables for customers on the same feeder, and again a characterisation of the associated uncertainties. Such data does not currently exist, and thus **a preliminary version of the tool must be created which does not fully adhere to rigorous Bayesian modelling**, but which captures the necessary data over a period of a few years (which is the role of Component 23 – and not required for the functioning of the tool), at which time the full Bayesian tool can be deployed.

4. The **tool will decide whether the demand of all customers may be aggregated together to form a single variable**, or whether they need to be aggregated as several groups, based on physical proximity on the network. In the latter case, the aggregated demand of each sub-group is modelled as a random variable, and the joint probability distribution of the set of variables must be modelled. This process will be dynamic, with e.g. warnings from monitoring data or customer complaints possibly leading to an additional subdivision of customers being created and a new joint distribution being modelled.

5. Data about network topology and component ratings for the specific network being modelled will be imported from the Spatial database, and used to construct a **model within a power system modelling environment**, for example IPSA, which has been used for demonstration purposes so far. Samples of carefully constructed 'plausible' simultaneous demands for each customer on the network are generated, and used in conjunction with the network model in the power system modelling environment to **construct simple modelled relationships between the aggregated demand variables and a subset of relevant network state variables**, e.g. the thermal loading of a circuit identified by the tool during an initial scan as potentially overloaded, or the voltage at the end of a particularly long circuit.

6. The modelled demand joint-distributions and modelled relationships between demand and the critical network state variables are combined to **produce** any **specified risk metrics**, the default being the average rate of occurrence, expressed as events per year, of a thermal loading or voltage violation. It is these outputs that would be compared against the NPg design criteria to establish if an existing network or a new design is acceptable or not.

2.2 The Main Algorithms

The operation of the tool is made possible through the set of algorithms presented in summarised form below. Our main concern here is the description of the algorithmic requirements of the first version of the tool, i.e. the Minimum Viable Product (MVP), although it also makes sense to refer to the ways in which those requirements will expand as the sophistication and scope of the tool grows through multiple iterations. It is important that the MVP is designed to allow such future flexibility. The main algorithms in the MVP must achieve the following:

1. Coordinate load-flow studies on the LV networks (where the load flows are solved by an external load flow engine), in order to (i) initially test whether circuits are likely to experience overloading or voltage violations, (ii) construct relationships between aggregated demand variables – either for all customers, or some subset of them - and specific network voltages and currents.
2. Construct a representation of what types of customers are connected to the end of each service cable. The simplest possible version is to assume all customers are either generic domestic customers (i.e. Elexon's 'Domestic Unconstrained') or street furniture, and that one or (in the case of looped services) more such customer exists at the end of each service cable. The more sophisticated approach, which is also included in the MVP, is to acknowledge the existence of non-domestic customers and categorise these into 2 or 3 types. In future versions, domestic customers could be categorised according to the house types, the presence of electric vehicles, fuel used for heating, and possibly socio-economic categories.
3. Generate samples of plausible simultaneous demands for each section of mains cable, based on the customer type representation described in paragraph 2 above. If it is found, after being run through the network model in IPSA (or alternative tool), that none – or very few – of the demand samples resulted in any thermal loading or voltage violations, then they must be scaled-up until this is the case. If the re-scaling means that most of the samples lead to a violation, then the re-scaling factor was too large, and a smaller one must be trialled. The generated samples exist for two purposes:
    a. To identify whether any nodes or circuit sections experience violations after only modest re-scaling of demands – if so, the LV network is identified as being 'at risk' of being overloaded or of delivering low voltage and requiring action. Further, the quantities that exceeded their limits after only small re-scaling are identified as critical network state variables.
    b. To enable the algorithm described in paragraph 4 to model the relationship between suitably aggregated demands and each identified critical network state variable. Crucially, these relationships must be modelled for the demand range surrounding the point at which violation occurs, so the sample generation algorithm must find a suitable full set of re-scaling factors to enable this.
4. Build models of the relationship between aggregated demands and the identified critical network state variables (i.e. voltages and thermal utilisation levels), by running the demand samples described above through the network model in the AC load-flow modelling environment. For the MVP, these models will take the form of simple linear and/or quadratic relationships, and will involve either the aggregated demand of every customer on the feeder, or for each phase of the feeder. Future version s of the tool will include an algorithm that identifies which demands should be grouped together and aggregated as the best explanatory variables. The algorithm will also examine whether only one aggregated demand variable is sufficient for an accurate model, or whether multiple explanatory

demand variables are necessary – with R-squared values probably being the most relevant metric, and the threshold impossible to set without first gaining insight from the data. The simplest (but not fastest) version of his will inevitably involve running through many possible demand aggregations, which may be quite computationally expensive. Future versions could introduce machine learning methods such as neural networks to predict which variables should be grouped together, based on learning from topographically similar networks as a training dataset.

5. Take the representation of customers from paragraph 2, the optimal grouping information from paragraph 4 and an expanding database of historical demand time series to produce a probability distribution for the aggregated demand at key points in the network, or if deemed necessary by (paragraph 4), a joint probability distribution for the set of aggregated demands. The algorithm will eventually work within a Bayesian framework, and the distributions produced will be known as prior distributions within this framework. For the MVP, if the number of domestic customers to be modelled, N, has been previously modelled by the tool, the distribution parameter already calculated can be reused. Otherwise, they can be calculated through the time series sampling-based method described in the Novel Analysis Techniques report. In later versions, computational expense can again be reduced through prediction by a machine learning algorithm trained on similar N values, or sets of N values in the multivariate case. If the actual penetration level of certain LCTs, including PV, on a specific network or feeder become significant, then it would be sensible to assume that they can be represented by an average PV generation profile for the historical period matched to the CLNR TC1a dates, and include them in the time series sampling method.

6. Extract and process (e.g. for data quality) all available demand-related data held in relation to the specific LV network, including smart meter data extraction from the Siemens EIP environment, and LV monitoring repository. In the ideal situation, and where the required aggregation number is allowed by regulations, the smart meter data should take the form of time series data. For the MVP, the distributions of load data must take the form of sets of histograms (e.g. one per season). This data will eventually be used to update the distributions, or joint distributions where necessary, to form posterior distributions in the Bayesian framework.

7. Use the modelled relationships between aggregated demands and critical network state variables, along with the updated probability distributions for demand to generate probability distributions for those network variables (thermal loading, voltage). Once these are established, they can be used to extract the desired risk-based metrics, as set by DNO managers.

2.3 Existing NPg Systems

The MVP is required to interface with or be presented with data from other NPg systems:

- **eAM Spatial** – There are two existing Oracle databases used to store asset information: Spatial is used primarily to store asset location data and acts as the geographic information system along with the web application iSMART which is used to view the data. eAM is the central asset database. In the MVP we only need to access the Spatial database. The Spatial database contains enough asset information for the MVP to work with. Future versions of the tool may require additional information stored in eAM such as an embedded generator's kVA size. eAM Spatial is located on NPg's Corporate IT network.

- **Siemens EIP** – This is the head end system that interfaces with the Data Communication Company's (DCC) smart meter gateway. It is this system that NPg use to request consumption data from smart meters. It is currently located on a dedicated secure environment with no physical network connections to other NPg networks. The MVP will interface with it by requesting aggregated half-hourly consumption information (per feeder) on a quarterly basis. The means of transmission of data into and out of the secure environment will be by CD-ROM/DVD-ROM.

The MVP also requires the following data:

- **Non-half hourly domestic customer annual kWh consumption** – These are currently collected by NPg and can be made available to the MVP with a manual upload every year.
- **CLNR half-hourly consumption readings** – The CLNR dataset TC1a contains approximately two and a half years of half-hourly load data for 8000 domestic customers in total – about half of which were filtered out to ensure a very high % of data present. This will be made available to the MVP by a one-off manual upload. There are also much smaller (and shorter) datasets capturing the consumption patterns of electric vehicles (EV) and air source heat pumps (HPs) that are used to represent these technologies' behaviour. Dataset TC6 contains metered data for 144 EVs, while TC3 contains metered data for 89 HPs. These – or possibly more recent data sets, can be used to form slightly crude average profiles for EVs and HPs, that can be included in the sampling process.

2.4 Requirements

2.4.1 User Requirements for MVP

The basic requirements of the first version of the tool, i.e. the MVP, from the perspective of the network designers, are the following:

- It must be easily understandable and intuitive. Indeed, the tool should be usable with only familiarisation training and without LV designers having to refer to documentation for carrying out common tasks.
- The tool should be relatively quick to run, consistent with volume of use case in daily workload. Any delay in carrying out a calculation shall be proportional to the frequency upon which the associated use case is executed. It should not involve any repetitive manual tasks.
- The front-end of the tool shall be an intranet web-based tool. This would enable quick access, be user friendly, and reflect the trend in software applications moving towards becoming cloud-based.
- The central purpose of the tool is to confirm or deny whether voltages and cable/ line utilisation levels are within the risk level deemed acceptable. However, the network designers shall not be required to understand the underlying statistical theory and methodology to make design decisions.
- If we express the expected frequency of occurrence as '1-in-*n* years', then we have *n* = 1/(exceedance expectation), where exceedance expectation is the average number of exceedance events per year, i.e. the number of half hour periods per year where the average (over the half hour period) voltage drop/ rise, or utilisation level, is greater than an agreed threshold. Calculating these exceedance expectation values is the main objective of the tool's modelling components, with the values inverted in a final step before presentation. For example, a result might say "the expected thermal exceedance frequency for cable 'x' is 1-in-11.43 years". The precise meaning of this statistical metric is that there would be an average of one half-hour every 11.43 years where the thermal rating of the cable is exceeded, if the network were to serve the same demand patterns repeatedly for many 1000s of years. For the avoidance of doubt, we note that this approach calculates voltage drops on the LV network, and the absolute voltage is based on the nominal HV voltage at the secondary distribution substation.
- The default threshold for exceedance is 100% - i.e. an exceedance event is where the line/cable utilisation is above 100% of its rated value, or the voltage at a node is above or below a permitted limit. However, the user will be able to calculate expected frequencies of occurrence for a different exceedance threshold on an ad-hoc basis for individual feeders. As a guide to acceptable expected frequencies of occurrence, the ACE 49 report recommendation, and the standards based upon it, seem to set 1-in-10 years as the acceptable level, although this is somewhat open to interpretation, and is not considered explicitly by designers or managers responsible for such policies.

- An unbalanced load-flow algorithm is required to calculate the relationships between demands, power flows and voltages. This will either be a licensed engine (e.g. IPSA), or open source (e.g. OpenDSS).
- The tool will deal primarily with the 'base-case scenario' for each feeder, representing NPg's best guess of the nature of demand patterns currently present on the system, along with three scenarios for unknown phase connections. These scenarios will be: perfect phase balance, moderately bad phase imbalance, and very severe phase imbalance. The base-case scenarios also correspond to the normal running arrangement for the feeder, i.e. where normally open points are open etc.
- The tool will calculate the necessary base-case exceedance expectation values for every LV feeder in NPg's networks as soon as the necessary network data has been imported from Spatial, so that they can be accessed very quickly by the tool's users.
- Users will be able to bring up the base-case scenario as an annotated diagram on their screen, and then be able to make edits, such as:
    - increasing existing demands at existing customer connections,
    - adding new demand to existing connections – particularly adding LCT demands,
    - adding new customer connections,
    - installing new LV assets in an existing LV network,
    - replacing existing LV assets in an existing network,
    - reconfiguring existing assets on an existing LV network,
    - creating a new LV network.
- If the expected frequency of exceedance occurrence is higher than the acceptable limit for some line/ cable utilisation or node voltage, in the MVP the designer will be able to design alterations to the LV network (see above) and re-calculate the expected frequency of exceedance occurrence, to test if the reinforcement would be sufficient. In later versions they will also be able to or deploy other solutions such as Demand Side Response or voltage control. The models for these hypothetical alternative versions of LV networks, associated with the designer who created them, will be stored as study scenarios separately from models of the network as it currently exists. These study scenarios should be easy to incorporate in to the existing model once associated relevant changes had been made to the as installed LV network.
- The ability to calculate earth loop impedance since the existing LV design tools currently in use do this.
- The ability to mode secondary distribution transformers – the HV voltage will nominally be set to 1 p.u. for the MVP, and calculated voltages are relative to this.

2.4.2 Future Extensions

The MPV should be developed so as to facilitate the incorporation of the following fundamental feature: in future versions:

- Users shall be able to be assigned to one or more groups, where groups can be created for different teams e.g. LV Designers Yorkshire.
- The fundamental method of producing probability distributions of aggregate demand – for a given number of customers, where combinations of energy consumption, smart meter and possibly LV network monitoring data are available – will need to be fundamentally different in versions after the MVP. For the MVP this is based on sampling from the CLNR dataset, but in future versions it will be

done more analytically, through Bayesian updating, once sufficient data becomes available for the construction of Bayesian prior parameters for joint distributions of aggregate demands.

2.4.3 Exclusions

This subsection outlines what the tool will *not* be able to do at the MVP stage – both from the perspective of users, and those maintaining the system. The most pertinent exclusions are:

- Modelling transient analysis, stability, harmonics, or disturbing loads / EREC P28 studies.
- Modelling multi-voltage network models (HV/EHV).
- Modelling meshed networks, i.e. radial networks only - which is adequate for NPg's needs.
- Automatically suggesting network design interventions, to resolve identified issues, whether conventional or smart.
- Modelling LV networks outside NPg's North-East and Yorkshire licence areas. The first release needs to be shown to be effective within NPg's licence areas before the tool is 'sold' to other DNOs. One exception would be if the funding model for the tool development involves funding from other DNOs.

2.4.4 Maintainability and Support

The application code for the main front-end should be written in a conventional programming language such as Python, Java EE, or C#.NET. Statistical algorithms may be developed in specialist software such as R, although Python is likely to be equally capable. It should not require any third-party software licences with the exception of the load flow engine; it will need assessing whether there are benefits to a commercial load flow engine such as IPSA over an open source product such as OpenDSS.

2.4.5 Security

The application source code for the core tool functions including a reference LV network shall be open source. A 'Release Manager' shall be appointed to make the source code available to the wider community. The Release Manager shall be responsible to ensure that no data specific to NPg's network is present in the source code. Components directly related to accessing NPg's internal data sources shall remain private and the source code not made freely available. Authentication/ authorisation for the tool will be linked into user's network login.

2.4.6 Legal Requirements

Access to and treatment of smart meter data shall comply with the NPg's Data Privacy Plan, SLC10A and GDPR requirements. Treatment of specific domestic consumer data shall comply with GDPR.

3 Case Study Feeder

Before getting to the detailed presentation of how the proposed tool will be divided into a number of components, and the precise functions of those components, we present here a case study feeder that will be used consistently to illustrate those functions. The chosen circuit is Feeder 7 from the Cranwood LV network, part of NPg's Yorkshire area distribution network. (To be completely precise, it is the feeder connected to Way 7 coming from the transformer, but there are only 6 feeders, with Way 2 not connected to any customers). A diagram of the feeder is presented in Figure 3.1 below, where the exact location of the roughly evenly distributed individual customers has been omitted, for ease of interpretation. Instead, the number of customers connected to each phase is indicated for each mains cable segment. The benefit of this representation becomes more obvious later in this document, when we consider future scenarios where EVs and photovoltaic generators (PVs) have connected to the network.

The case studies provided in the Novel Analysis Techniques report[2] were restricted, for clarity and brevity, to looking at the utilisation of the first branch of each feeder supplied from the secondary distribution transformer and the voltage at a single node at the end of each feeder. With this case study, we do not maintain this restriction, and explore what issues the tool must deal with when we examine other sections of the feeder. We therefore assume that we want to analyse the node voltage at all mains cable ends – there are nine of these in our case study, and they are labelled A-H, in Figure 3.1.

We also want to monitor the branch utilisation (i.e. power flow normalised by the rated value) at the transformer, the link boxes, and at every mains joint (not service cable joints) where the conductor specification changes, i.e. where the network tapers. There are 15 such junctions on our chosen feeder, labelled 1 – 15 in Figure 3.1. Where the joint is simply a change in cable specification, the utilisation variable in question is naturally given by the power flow divided by the smallest of the two ratings. At joints where three or more cables are jointed together (points labelled 5, 6, 12, in Figure 3.1) we need to know the utilisation on each of the cables, and each needs to be modelled as utilisation variables.

This particular feeder provides a rather rich case study due to having several sub-branches and tapered conductor sizes. It is comprised entirely of cables and has two link boxes connecting it to the Crandyke 1 feeder and Crandyke 4 feeder. Currently, there are 116 single-phase domestic customers connected, and one three phase customer. This shall be referred to as the base case scenario for the feeder, and will be used exclusively when explaining the component functions. In Section 5, several alternative scenarios involving load growth on the feeder will be presented, accompanied by an explanation of how the tool would calculate the impacts of this growth on the network.

While both the network specifications and basic description of the customers in the base case scenario correspond precisely to the real network, we do not restrict ourselves to real patterns of demand arising from those customers, and the overloading problems found by the tool in the case study narrative are created to maximise illustration of the tool's functioning. The same goes for the number of customers with smart meters and the presumed quality of the data they record.

[2] Available at: https://www.northernpowergrid.com/asset/0/document/4918.pdf

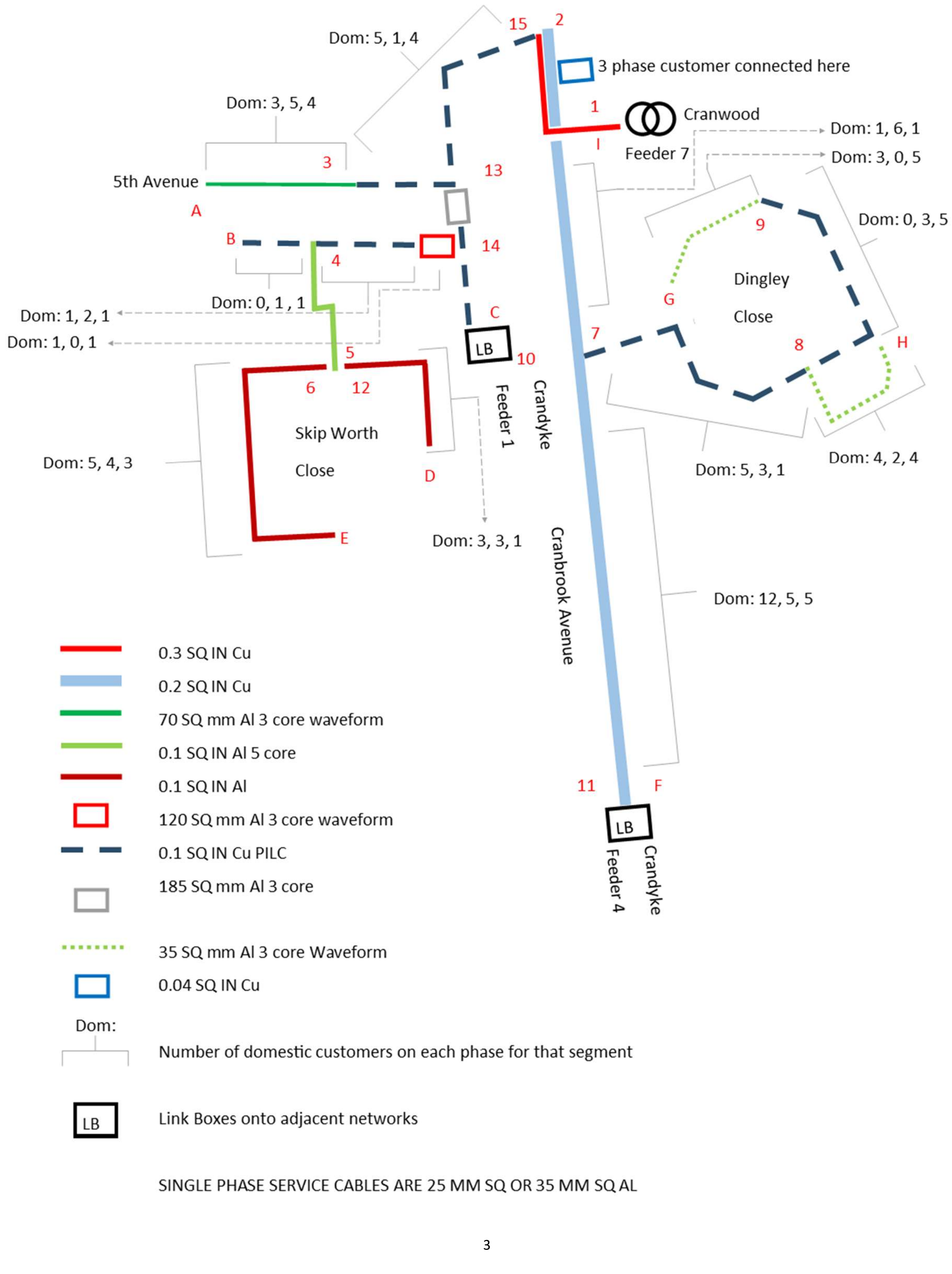

Figure 3.1. The Cranwood Feeder 7 case study, base-case scenario

3

Following the identification of all network variables, the fundamental task of the tool is to identify which subset of variables i.e. those relating to voltages / utilisation at the key network nodes identified in Fig 3.1– if any – have probability distributions such that the probability of exceeding their safe/ statutory limits are not negligibly small (e.g. might exceed limits with a frequency of roughly in 1-in-10 years), store the results and present them to network designers on request.

Due to the presence of the two link boxes, there exists several different possible running arrangements for the feeder, and these are presented in Table 3.1. below. For future versions of the tool, we expect it to automatically identify these, and calculate base-case exceedance expectation values for each operational scenario. However, the requirement for the MVP is simply to calculate these for the normal running arrangement, i.e. in this case with both link boxes. open.

Running arrangement	**Link box status**
1	Normal arrangement: Both link-boxes open, Feeder 7 fed from Cranwood.
2	Alternate running arrangement: Link-box to Crandyke Feeder 1 closed, Cranwood Feeder 7 and Crandyke Feeder 1 fed from Cranwood, link-box to Crandyke Feeder 4 open.
3	Alternate running arrangement: Link-box to Crandyke Feeder 1 closed, Cranwood Feeder 7 and Crandyke Feeder 1 fed from Crandyke, link-box to Crandyke Feeder 4 open.
4	Alternate running arrangement: Link-box to Crandyke Feeder 4 closed, Cranwood Feeder 7 and Crandyke Feeder 4 fed from Cranwood, link-box to Crandyke Feeder 1 open.
5	Alternate running arrangement: Link-box to Crandyke Feeder 4 closed, Cranwood Feeder 7 and Crandyke Feeder 4 fed from Crandyke, link-box to Crandyke Feeder 1 open.

Table 3.1. List of running arrangements for Cranwood Feeder 7.

4 Functional Specification

This section presents the detailed functional specifications for each component of the tool presented here, including all the details of how the components will interact. The specifications are given for the MVP, but there is a section describing envisaged future developments for each component. Components such as the load flow engine and smart meter data cleanser could be developed elsewhere or be off-the-shelf components. These other components and databases may also be used by other design tools.

4.1 Architecture/ Component Overview

The tool is conceived as comprising a number of components, along with the set of connections between them, passing data and triggers to initiate specific actions. The components can be categorised as two types: (i) scripts, programs or applications; (ii) databases, or collections of tables within an existing database. Figure 4.1 below presents a schematic diagram of all components, and their connections, with components of type (i) represented by yellow icons, and type (ii) with blue icons. Data sets stored on physical media for entering into the tool are represented by grey CD icons.

It can be seen in Figure 4.1 that the tool's 'ecosystem' comprises of the following entities:

(i) the secure smart meter data storage environment;

(ii) data sets that are inputted into the system manually on an irregular basis, such as network monitoring data;

(iii) the standalone *smart grid design tool enterprise server* – all of the interconnected component that accept network and customer data, and convert them into network overloading risks, to be viewed by network designers;

(iv) NPg's corporate IT Network, which will include previously existing relevant components, most notably Spatial, along new components related to the tool;

(v) an enterprise bus, facilitating and regulating the flow of data between the design tool enterprise server and the corporate IT network;

(vi) web browsers on network designer's personal computers, will be their means of interacting with the system.

The full descriptive names for each numbered component are presented below:

1. Spatial Data Extractor and Network Model Constructor
2. Database of Network Models, Compatible with the Load Flow Engine
3. Load Flow Engine Module
4. Voltage and Utilisation Variable Selector
5. Demand Variable-Set Constructor
6. Network Response Characteriser
7. Load Flow Engine Operator
8. Smart Meter Data Requestor
9. Data Extractor for Siemens EIP Environment
10. Smart Meter Data Integration Module
11. Annual Energy Consumption Data Uploader
12. Database for Annual Consumption and Smart Meter Data
13. Metered Non-Domestic Demand Uploader for Half-Hourly Non-Smart Meter Consumption data
14. Store for Intermediate Model-Building Data
15. Demand Variable-Set Sampler
16. CLNR Consumption Time Series Store
17. Probability Distribution Fitter & Exceedance Expectation Calculator

18. Exceedance Expectations Store
19. Smart LV Design Web Application
20. Network Designer User Interface (Web Browser)
21. Smart LV Design Database
22. Ordnance Survey Master Map
23. Joint and Marginal Probability Distribution Learning Database

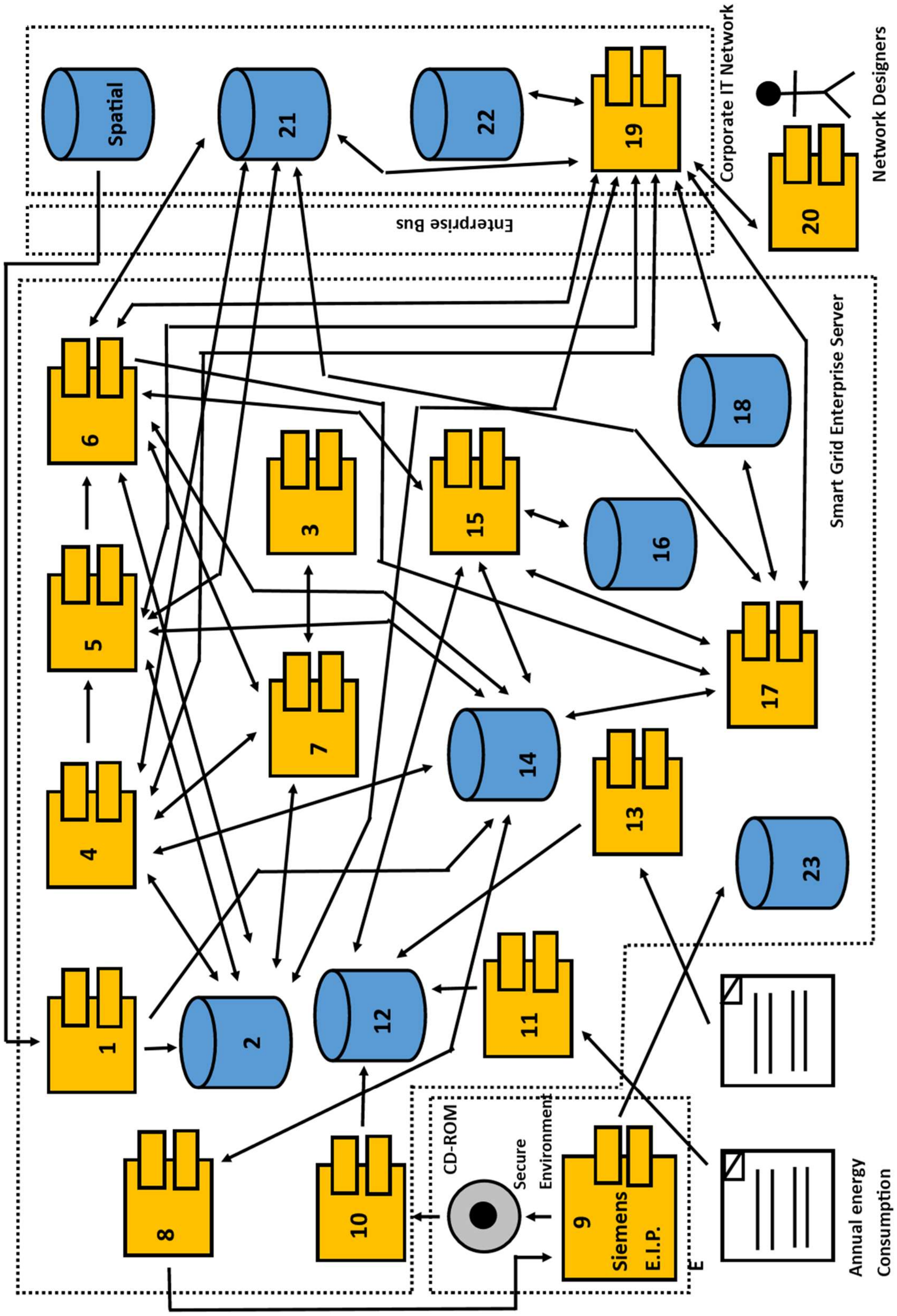

Figure 4.1. All components and connections between them, comprising the Smart LV Network Design Tool.

The following subsections of this report provide a detailed and structured presentation of the components.

4.2 Component 1: Spatial Data Extractor and Network Model Constructor

Functionality

- Imports network data from the Spatial database located on NPg's Corporate IT Network, and uses it to create electrical connectivity models of LV networks in a form that's ready to be run on the load flow engine software, Component 3, and saved in Component 2. An alternative way of expressing this is that it synchronises the data stored in the Spatial database with Component 2 located within the Smart Grid Enterprise Server.
- The resulting data models will represent electrical connectivity and asset electrical data (circuits, transformers, and are likely to be based on the Common Information Model (CIM).
- The data models will also contain sufficient Geographic Information System (GIS) data to be able to display the location of circuits and customers on background Ordnance Survey maps in the web application.
- The resource required to develop this component may be shared with other projects, as Component 2 (the Database of Network Models) will also be used by other tools in deployed in the Smart Grid Enterprise Server and not just the tool described in this report (Smart LV Design).
- It also needs to capture customer load location (exit points) and phase connectivity, and data on the presence of LCTs connected to each exit point, where it is known to exist. Where the connection phases of customers are not known, for the first version of this tool, they should be randomly allocated. The MPAN of the customer(s) at each connection point must be part of the model. This customer connection data is sent to Component 2 and Component 14.
- The models would mainly be populated from the Spatial database, extracted using SQL queries and using an appropriate scripting language to automate periodically as the network connectivity changes.
- Relating this to our case study, the 'journey' of the Cranwood Feeder 7 case study through the tool – or more specifically the Smart Grid Enterprise Server, begins very simply with a model of the entire Cranwood LV network being constructed by this component, and being sent to Component 2 for storage.
- It is likely that a data administrator will be required to perform data cleansing on the exported data before loading it into Component 2, as the data will need to be fully audited and cleansed prior to being imported, ensuring that the full connectivity model is intact. Cleansing must include ensuring that all conducting types map to known cable types, and ensuring that all LV electrical components are associated with the correct secondary distribution substation.
- Arrangements will be required to validate the imported models to confirm that data cleansing issues have been resolved.
- Either the data administrator or some automated function within the tool or must ensure the models are up to date, i.e. models should be updated if any LV networks are modified in any way or any customer data changes e.g. the connection of a new LCT.
- The magnitude of the data cleansing / validation / updating activity should not be underestimated; at the moment models are created by design engineers as required, whereas this proposal is to automatically and routinely assess the entire LV network for unacceptable voltage and thermal violations.

Mode of operation

- Creates a physical and electrical network model upon receipt of the required cleansed data before automatically storing them in Component 2, as described above.

- Data being passed through the component will mainly be freshly cleansed data, which gradually expands the number of LV circuits represented by the tool, until it includes every LV network operated by NPg.
- Occasionally the data being processed will represent an LV network that is already represented by the tool, but where a physical or electrical change has occurred such as a reinforcement being implemented.

Event Sequences and Triggers

- Either runs as a batch process or triggered to operate when data is changed in Spatial.
- Sends a data writing request to Component 2.

Data Inputs

Network data held in the Spatial database, including MPANs.

Data Outputs

- Physical geographical and electrical network connectivity model stored in Component 2.

Developments in Future Versions

- For the MVP, only models for the network under normal running arrangements will be considered, i.e. normally open link boxes will only be modelled as open. However, all plausible running arrangements will be considered in future versions.
- In future versions of the tool, the uncertainty associated with phase connections could be made explicit, by essentially 'shuffling' the connectivity of customers with unknown phases between the 3 phase connectivity options for each run of the load flow engine (Component 3), to give a probabilistically weighted representation of the extent of phase imbalance. That would likely to be the responsibility of this Component, but could potentially be executed by Component 3.

4.3 Component 2: Database of Network Models

Functionality

- Stores the network models created by Component 1, and allows them to be read by the Load Flow Engine Module, Component 7.
- This would be an electrical connectivity network model possibly implemented in a GIS enabled database such as PostGIS. It would have a CIM compliant interface.
- This database may be used by NPg tools, other than the Smart LV Design, and as such the development effort may be shared with other projects.
- Records should contain a time stamp of when they were created or edited, and the component will allow Components 4, 5 and 6 to check which networks have been added or edited since their last batch operation.
- Relating this to our case study, the Cranwood network model created by Component 1 is stored here, waiting to be read by the module that coordinates load-flow studies, Component 7. This will occur when either of Components 4, 5 or 6 accesses the database and notice that the Cranwood network has been created (or edited) since the last data upload.

Mode of Operation

- Behaves as a straightforward database (i.e. no stored procedures etc), allowing read and write requests.

Event Sequences and Triggers

- Allows Component 1 to create new records and update existing ones.

Data Inputs

- Electrical connectivity models of the form accepted by the load-flow engine, with additional GIS markers to identify equipment/customer locations.

Data Outputs

- Same as Data Inputs

Developments in Future Versions

- In future versions of the tool, multiple running arrangements will be stored for each feeder.

4.4 Component 3: Load Flow Engine Module

Functionality

- Given a LV electrical network model, and a full set of customer demand values (from the sampling modules), it calculates the resulting current in branch and voltage at each node in the network.
- This is not a component that needs to be built, rather a product such as IPSA or OpenDSS can be deployed. To be more precise, it would need to comprise of several instances of these, in order to increase model throughput.
- The component is controlled by a dedicated module, Component 7, that will supply 1000 sets of customer demand values, allowing calculations to be run 1000 times, for each main feeder. Each set of demand values includes a sampled demand value for each MPAN on the feeder. This may be repeated for re-scaled demand value sets.
- In terms of the case study, this component will be supplied with the Cranwood LV network electrical model by Component 7, in turn supplied from Component 2. It will also be supplied (by Component 7) with 1000 sets of demand values for Cranwood Feeder 7, produced by Component 15 – with different demand values but the same network model for each of the 1000 calculations. This will also be conducted for all other main feeders in the Cranwood LV network. The resulting voltage and utilisation values at each node will be collected by Component 7, and passed on to either Component 4 or 6, depending on which aspect of network characterisation is being addressed, as will be explained bellow.

Mode of Operation

- Operate 1000 times with the same network model, but different sets of customer demand values, that are all samples of plausible peak-period customer demands that the network is expected to experience. It outputs a distinct and full set of voltage and utilisation values (i.e. at each node), for each set of customer demands.
  It will conduct an unbalanced-phase load flow and this document will assume this to be the case.

Event Sequences and Triggers

- The unit is triggered to execute load flow calculations by Component 7, using the data provided by that component, and returns to it a full set of network utilisation (circuit thermal loading) and voltage values.

Data Inputs

- An electrical network model containing nodes, branches and impedances of circuits.
- The secondary distribution transformer's electrical model and tap setting.
- A nominal HV voltage at the secondary distribution transformer.
- 1000 full sets of demand values, and assumed power factor, at each MPAN location.

Data Outputs

- The voltage at each node for each of the 1000 demand data sets
- The utilisation level in each branch, for each of the 1000 demand data sets expressed as raw Amps and kVA values, before normalisation into utilisation %.

Developments in Future Versions

Future versions will calculate, by default, extreme *minimum* values of demand minus generation – and from this, exceedance expectations for high voltage excursions beyond allowed limits. Currently, such values are calculated ad hoc if requested by the tool's user demand.

4.5 Component 4: Voltage and Utilisation Variable Selector

Functionality

- This component runs calculations to determine which, if any, of the circuits or nodes belonging to each main feeder being examined are loaded heavily enough, such that there's a reasonable chance that their thermal or voltage limit will be exceeded e.g. at least once every 10 years. If there is no such chance, for a given feeder, then that feeder is deemed safe from overload and excessive voltage excursions, and rigorous probabilistic modelling is not required. The subset of circuits and nodes identified as being at reasonable risk of occasional exceedances determine the set of output variables of interest – with one set for each feeder comprising the LV network. Specifically, the utilisation and voltage values at those risky nodes, branches, transformers and link boxes form the set of output variables. This identification of the 'relevant subset' of network data is considered to be *meta data* associated with the feeder, and will therefore be saved to the Store for Intermediate Model-Building Data, Component 14, or the meta data store.
- The component must therefore establish, through a computationally inexpensive and therefore approximate method, the risk of voltage or thermal utilisation violation for every node and branch etc in the feeder. The threshold risk level will be set by NPg, and will take the form of the expected frequency of occurrence being higher than 1-in-n years.
- This will be achieved by carefully selecting 1000 sets of demand values (each set comprising a demand value for each MPAN on the feeder) and running a load flow in Component 3 for each set (under the coordination of Component 7).
- The 1000 sets of demand values will be created by Component 15, under the instruction of this component, and the process is described in the Component 15 section.
- As a result of the 1000 load flows, there will be 1000 values for each of the network voltage and utilisation variables. A rule will be in place to determine if each variable has any sufficiently large instances (within their set of 1000 values) to classify it as 'potentially at risk' of overloading or an excessive voltage excursion. This rule might be e.g. at least one of the 1000 utilisation values is >= 90%, or at least one voltage excursion is >= 90% of its statutory maximum value. The appropriateness of any rule would have to be tested during the tool's development; with the intention being to minimise the volume of LV networks that need to be subject to more rigorous assessment.

- To develop the case study narrative, we need to imagine that when this component was investigated for Cranwood Feeder 7, it was found that 3 voltage variables and 3 utilisation variables were in breach of the 'potentially at risk' rule, e.g. 2 of the 1000 values for the thermal utilisation in branch 1 were > 90%. The voltage variables are at the ends of two branches: the red and yellow phases at node *E* and the red phase at node *G* (using the notation adopted in Figure 3.1), while the thermal utilisation variables are the 3 phases of the mains cable connected to junction 1, i.e. the first cable segment supplied from Way 7, which supplies all customers on the feeder.

Mode of Operation

- Operates as a batch process automatically at the beginning of each quarter, but can also be manually triggered by the tool's administrator, if many new LV networks have been synchronised in Component 2.
- It also operates on individual feeders, when a network designer is actively working on it. Typically, a designer will construct an alternative version of a feeder, i.e. a scenario that is yet to come true, e.g. where a significant take-up of LCT demands is being predicted, or the network has been extended to accommodate new properties. The designer needs to know initially whether any network variables have moved into the 'at risk' category due to the changes included in the scenario. In such cases, this component will be triggered into action by Component 19, the Smart LV Design Web Application, and will use data stored in the dedicated database for such hypothetical scenarios, the Smart LV Design Database Component 21.

Event Sequences and Triggers

- Batch operation automatically or manually triggered.
- Triggers Component 15 to generate demand value sets.
- Then triggers Component 7 to coordinate load-flow studies.
- When the batch process is complete, triggers a batch process in Component 5.
- Also triggered by Component 19 to perform analysis on case-study versions of individual networks.

Data Inputs

- None required – coordinates the generation and sharing of data between other components.

Data Outputs

- Following batch operation, will output meta-data stating which power flow and voltage variables require modelling for each feeder, for storing in the Store for Intermediate Model-Building Data or meta data store, Component 14.
- When triggered by Component 19, the output will be stored in Component 21.

Developments in Future Versions

- For the MVP, feeders will only be tested in their normal running arrangements, i.e. with link boxes open. However, all alternative configurations will be investigated in future versions.

4.6 Component 5: Demand Variable-Set Constructor

Functionality

- This component is relevant to feeders that have already been identified (by Component 4) as possessing at least one circuit or customer node where the power flow or voltage limit, respectively, is reasonably likely to be exceeded at least as often as e.g. once every 10 years (if this is the chosen policy). This is true for our case study.

- One of the most fundamental aspects of our overall modelling approach is we assume that each voltage and utilisation variable is given by a 2nd order polynomial regression model, i.e.

$$V = \alpha_0 + \sum_{i=1}^{i=m} \alpha_{1,i} d_i + \sum_{j=1}^{j=m} \alpha_{2,j} d_j^2 + \varepsilon \quad (1)$$

where V is a voltage or thermal utilisation variable, the α_0, $\alpha_{1,i}$, and $\alpha_{2,j}$ are the model coefficients for that variable, and the regression variables $d_1, \ldots, d_m$ are aggregate demand variables, and ε is the prediction error. The task of this component is to determine the exact nature of those aggregate demand variables for the best-fitting model, e.g. it might determine that when predicting the voltage at some node, setting d_1 to represent the aggregate demand of 18 domestic customers makes more accurate predictions than if d_1 represents the aggregate demand of e.g. 17 or 19 customers, with the most suitable metric for success probably being R-squared values.

- For the MVP, the task of this component will be made much simpler by restricting the model choice to ***only having one aggregate demand variable*** (i.e. $m = 1$ in the equation above, so we no longer need subscripts for d). We consider this to be reasonable based on case studies within the Novel Analysis Techniques report.
- Further, we assume that the single variable, in the case of *voltage regressors*, is given by: all demands on the same phase and on the same branch, where the branches in question are those created at the feeder's first splitting junction, i.e. the first splitting junction down-stream of the way. If there are a minority of customers connected upstream of the split, these will be included for both branches. If the majority of customers are upstream of the split, then the only aggregate demand used in relation to voltage calculations is the sum of all demands on the feeder.
- In the case of *thermal utilisation regressor* variables, it is simply the sum of *all downstream* demands, on the same phase.
- For our Cranwood Feeder 7 case study, it was identified by Component 4 that all three phases of the first cable segment (immediately downstream of Way 7) are worthy of modelling. The three corresponding aggregate variables are, according to the principle outlined above, comprised of 43, 35 and 36 domestic customers, respectively, with a single 3-phase customer split between them.
- Cranwood Feeder 7 has a splitting junction close to Way 7 into two branches: one of them branches further, with the branches ending at the nodes A, B, C, D and E; the other branches further, with the branches ending at nodes F, G and H. There are no customers upstream of the first split. The sets of domestic customers on branches ending up at A, B, C, D and E – on each phase – is [18, 16, 15], with no three phase customers. For the other initial branch, ending up at F, G and H, the set of customers on each phase is [25, 19, 21], with one three phase customer.
- The voltage variables identified by Component 4 as requiring modelling are the red and yellow phases at node *E*. According to the customer numbers above, and the MVP rule for voltage variable regressors, the aggregate demand variable for the red phase is 18 domestic customers, while for the yellow phase it is 16 domestic customers. The 3rd voltage variable identified by Component 4 was the red phase at point *G*, which consists of 25 customers.
- This model will clearly not predict the network variables as calculated by the load-flow engine perfectly, particularly voltages, due to the approximations made for simplicity for the MVP. Another factor with voltages in an unbalanced load-flow analysis is that the demand in the other two phases will likely have some impact on each phase's voltage, which is not considered for the MVP.

Mode of Operation

- One mode of operation is batch calculations, working through all LV network feeders that have been added or edited since the last batch operation, as indicated by time stamps in Component 2.
- The other mode is operation on single feeder case-studies, created by network designers, analogously to Component 4.

Event Sequences and Triggers

- Batch operation is triggered by Component 4, when that component has finished its own batch operations.
- Triggers Component 6 to start batch operations when finished.
- Operation on single case studies are triggered by Component 19, analogously to Component 4.

Data Inputs

- Data on the variables requiring investigation for each feeder, from the as-built network Store for Intermediate Model Building Data, Component 14, or the database for case studies, Component 21.
- The network model, from ether the as-built network store, Component 2, or the database for case studies, Component 21.

Data Outputs

- Demand variable meta data, to be stored in Component 14 or Component 21, depending on the mode of operation, including the MPAN numbers associated with each network variable to be modelled, for each feeder.

Developments in Future Versions

- In future versions of the tool, the rather strong assumptions – that the regression model involves only one aggregate demand variable, and that it is the total upstream (for voltage) or downstream (for utilisation) demand – will not be maintained. That means that this component must construct a number of plausible models, differing in the construction of the aggregate demand variables, possibly where some models have multiple demand variables, while others do not. For each variant of the aggregate demand variable(s), this component must instruct Component 6 to fit the regression model, and establish the best demand variables by comparing their performance (in terms of an appropriate metric such as the adjusted R^2).
- An algorithm would have to be developed with the ability to make 'informed guesses' about suitable candidate variables, and it is anticipated that this would be based on some form of unsupervised clustering.

4.7 Component 6: Network Response Characteriser

Functionality

- This component's function is to find the optimal model coefficients for the regression models representing relationships between aggregated demand variables and network variables, i.e. voltages and power flows, as represented by the equation in the previous section.
- These will be used to convert probability distributions for the aggregated demand variables into probability distributions for voltage and thermal utilisation values.
- For each feeder identified as requiring rigorous modelling, and then for each network variable identified as essential to model, this component is supplied with datasets comprised of the aggregate

demand input and corresponding network variable value, as calculated by the load flow engine (Component 3).

- This component will instruct Component 15 to produce 1000 sets of demand values, where each set has one demand value for each MPAN, in exactly the same was as Component 4, and sends them to the load flow operator, Component 7, to obtain corresponding network variable values. It learns which MPAN numbers form the aggregate variable (which was calculated by Component 5) by extracting this information from either the as-built network meta data store, Component 14, or the database for case studies, Component 21, as appropriate. By aggregating the correct subset of demands into a single variable, we have corresponding sets of independent and dependent variables for the regression.
- However, this component will also multiply all the data by 2 or 3 different scaling factors, and re-send to Component 7, with the scaling factors calculated so as to ensure a wider range of input and output values, and to ensure that there are a significant portion of the output values on both sides of the 100% utilisation threshold, or the statutory voltage limits, and without straying too far from those thresholds. There should be an option to managers to re-set these thresholds as e.g. 110% utilisation.
- As was discussed in the previous component's section, for our case study there were 6 network variables identified as being potentially at risk of occasional limit exceedance, and therefore this component would have to fit 6 regression models.

Mode of Operation

- Batch operation, analogous to Components 4 and 5, working through new or edited feeders. When its batch operation is complete, i.e. all identified variables are modelled for each LV network recognised by the tool, it triggers Component 17 to begin its batch operation, also working through each variable.
- Also analogously to 4 and 5, it will operate responsively on a single feeder when a network designer wishes to test a scenario.

Event Sequences and Triggers

- Triggered to start batch operation by Component 5, when it finishes its own batch operations.
- It is also triggered by Component 19 to perform on case-study versions of single feeders.
- Triggers Component 7 to coordinate load-flow studies.
- Triggers Component 17 when batch operation is over.

Data Inputs

- Extracts meta data from Components 14 or 21, instructs Components 7 and 15 to generate data for itself

Data Outputs

- Regression model coefficients, sent to either Component 14 or Component 21, depending on the mode of operation.

Developments in Future Versions

- In future versions, Component 6 could be used by Component 5 to determine the optimal set of aggregated demand variables, as previously described.
- The models fitted could be made more sophisticated, e.g. with the introduction of interaction terms among the explanatory variables, although this will not be a very high priority.

4.8 Component 7: Load Flow Engine Operator

Functionality

- This component coordinates Component 3 to run 1000 load flow studies, all on the same feeder, but with a different set of demand values for each run, where each set includes a demand value for each MPAN on the feeder. Other data supplied to the engine that remains fixed across the 1000 runs are: the secondary distribution transformer's electrical model and tap setting; the HV voltage at the secondary distribution transformer; and the Link box locations and usual state.
- The component is the middle level in two hierarchical structures: Component 4 → Component 7 → Component 3, and Component 6 → Component 7 → Component 3, where the purpose is simply to allow components 4 and 6, respectively, to operate.

Mode of Operation

- Acts once when triggered by Component 4 or Component 6, which means it will effectively act as a batch process when those triggering components are in batch operation. It will be triggered on a one-off basis (by the same components) when a designer creates and tests a new scenario for a feeder.

Event Sequences and Triggers

- Triggered to operate once by either Component 4 or Component 6.
- Triggers Component 3 to operate 1000 times (supplying it with a complete set of inputs each time).

Data Inputs

- The network related data, which this component will supply to the load flow engine (Component 3) unchanged 1000 times. This consists of:
  - An electrical network model containing nodes, branches and impedances of components.
  - The secondary distribution transformer's electrical model and tap setting.
  - The HV voltage at the secondary distribution transformer.
  - Link box locations and usual state.
  - Knowledge of which column in the demand data matrix corresponds to which location in the network.
- A matrix of demand values. The number of columns will be equal to the number of MPANs on the feeder being modelled, and there will be 1000 rows.
- When triggered by Component 6, it will also need a list of which subset of output variables from the load flow engine (i.e. active power and voltage values) needs to be captured and exported. When operated by Component 4, all current and voltage values must be captured and exported.

Data Outputs

- Matrices of current and voltage values, with one column for each active power, reactive power and voltage variable that needs to be captured (all such variables when triggered by Component 4) and 1000 rows – one for each run of the load flow engine (Component 3).
- For our case study feeder, this means that when triggered (and supplied with data) by Component 4, the output is a 116 x 1000 matrix, but is a 6 x 1000 matrix when triggered by Component 6.

Developments in Future Versions

- In any future version where the uncertainty surrounding phase connectivity is reflected by randomly 'shuffling' demands between phases, then a matrix keeping track of the phase connectivity of each MPAN during each of the 1000 runs will have to be added to the inputs, and the operation of this component might be slightly more complex.

4.9 Component 8: Smart Meter Data Requestor

Functionality

- The objective of this component is to calculate what smart meter consumption data the tool wishes to extract from the Siemens EIP, and write that specification onto a CD-ROM, to be sent to the Siemens EIP located in the secure smart meter environment. As there is currently no network connectivity into the secure smart meter environment all data into and out of the environment has to be transmitted on physical media.
- The full request will include the data wanted for each main feeder of each LV circuit stored and modelled by the tool – or more specifically, the design tool enterprise server. For the MVP, there will be precisely one dataset requested per feeder, with the aspiration of increasing this in later versions. The requests are made quarterly, at the end of each season, and are for smart meter data recorded during the most recent season. For example, a request will be made in early March for winter season data (December to February).
- The default data requested, per feeder, is an aggregated consumption time series, with 30-minute resolution and covering 3 months, for a selected subset of customers on the feeder. That is, this tool specifies which customers are in the chosen subset for a given feeder – as identified by their MPAN, and if any of the customers in that subset have high-quality smart meter consumption data, then their consumption time series will be added to the aggregate series.
- If the number of customers in the subset with high-quality data is smaller than the minimum allowed by the regulator due to privacy protection, currently assumed to be 2 or 3, then the aggregate time series will be converted to a histogram, as described in the Component 9 section below.
- The subset calculated by this component are those customers that are common to as many selected aggregate demand variables (as identified by Component 5) as possible. Imagine that, for some specific feeder, Component 4 identified that there are N network variables that require probabilistic modelling, and that Component 5, accordingly, has identified N groups of upstream or downstream customers, for which the aggregate demand patterns must be calculated. Imagine further that there exists a subset $\boldsymbol{S_N}$ consisting of s_N customers that form part of all N aggregate variables, while subset $\boldsymbol{S_{N-1}}$, consisting of s_{N-1} customers, is the largest group of customers that form part of *N-1* aggregate variables, and so on down to the largest group that only appear in one aggregate variable. The tool calculates the sums: $N \cdot s_N,\ (N-1) \cdot s_{N-1},\ \ldots, 1 \cdot s_1$ and notes the largest. If the largest sum is $(N-m) \cdot s_{N-m}$, then $\boldsymbol{S_{N-m}}$ is the subset of customers for which data is requested by this component.
- This process can be demonstrated through the example of our Cranwood case study feeder. As previously noted, Component 4 determined that 3 thermal utilisation variables and 3 voltage variables required probabilistic modelling. Then, Component 5 determined that the aggregate demand variable used in the regression equation for the three utilisation variables consists of 43, 35 and 36 domestic customers, respectively (temporarily ignoring the 3-phase customer). Since these variables relate to different phases, they are all different customers, and there is therefore no subset of customers common to all of them.
- Indeed, since we have 3 variables that relate to the yellow phase, two that relate to the red phase and one that relates to the blue phase, the best that can be hoped for is a subset common to the

three yellow phase variables, i.e. S_3. However, we saw in the Component 5 text that the yellow phase domestic customers were split into two mutually exclusive groups, one group used in the regression equation for the voltage at node E, and the other in the regression equation relating to node G. This means that the best choice must be S_2 or S_1 (i.e. subsets that are common to two aggregate demand variables or only a single variable).

- Considering S_2, we have a set of 25 customers that are common to the red phase voltage at node *G* and the red phase utilisation variable. This is bigger than the 19 customers that are common to the yellow phase voltage at node *G* and the red phase utilisation variable. The relevant value associated with S_2 is therefore $2 \cdot 25 = 50$. The biggest group found only in one aggregate demand is the 43 customers associated with the red phase utilisation variable, and since this is smaller than 50, the S_2 group wins. In other words, this component will calculate (and write in the CD-ROM) that for Cranwood Feeder 7, it wishes to extract smart meter consumption data for the 25 domestic customers identified by Component 5 (and stored in Component 14) that are aggregated together to form the regression variable associated with the red-phase voltage at node *G*. This process is repeated for each feeder that is represented by the tool, and shortly after the end of each quarter, the complete set of requests is sent to the Siemens EIP environment.

Mode of Operation

Batch operation, working through each feeder, shortly after each quarter (with the first quarter being Mar – May).

Event Sequences and Triggers

Automatically triggered into calculating the desired customer subsets by the date. Writes results onto CD-ROM when manually triggered.

Data Inputs

The MPAN numbers associated with each network variable to be modelled, for each feeder represented by the system, extracted from Component 14.

Data Outputs

Sets of MPAN numbers to be requested from the Siemens EIP environment, one for each feeder, passed to Component 9 within the Siemens EIP isolated environment via CD-ROM.

Developments in Future Versions

In future versions it will be possible to increase complexity by requesting data for more than one set of MPANs per feeder.

4.10 Component 9: Data Extractor within Siemens EIP Environment

Functionality

- This component queries the Siemens EIP smart meter database, to extract all available 30-minute consumption time series, covering the most recent 3-month period, for a carefully selected subsets of customers, represented by their MPANs. The MPANs required for each feeder is provided on a CD-ROM produced by Component 8.
- Upon reading an MPAN, the first response for this component is to check whether the customer had a smart meter installed, commissioned and accessible by NPg for the entire 3-month period. If true, it will then make an assessment of quality – initially screening and removing clearly erroneous data, such as infeasibly large values, then analysing the amount of missing or erroneous data. If the data quality passes a number of quality control criteria (not yet determined, but involving both the frequency and duration of periods with missing or erroneous data), then the MPAN will be accepted

as a valid entry for the 3-month period. Any missing or erroneous data will be 'filled in', either by interpolation for very short intervals of missing data, or with mean values for the season and time of day, for longer intervals.

- Once all of the available valid entries have been established and 'filled-in', for the subset of MPANs requested for a particular feeder, they will be added together to form a single time series. If the number of valid entries is equal or higher than the regulated minimum (for an aggregated series that can be viewed by DNOs, for privacy protection), this series is the component's final output for the feeder in question. This output will be written onto a CD-ROM, to be sent to Component 10, the smart meter store.
- If, however the number of valid entries is smaller than the regulated minimum, the time series will be converted into a set of histograms – one for each half-hour time of day, and these will be written onto the CD-ROM (if permitted by the DPP). In both cases, the MPAN numbers of each contributing customer is also written onto the CD-ROM, along with the total % of valid data that was used to create the series (as opposed to filled-in data). The 'regulated minimum' here refers to the tool's need to be compliant with NPg's data privacy plan (DPP), and the SLC10A legislation.
- In this activity, for every feeder, this component gathers and fills-in every valid time series, i.e. it looks for every MPAN associated with the feeder. If the total number of valid series is greater or equal to twice the regulated minimum number of aggregated customers, then the series will be split into two groups – randomly, but where neither side has fewer than the regulated minimum number. The series within each group will then be aggregated and both resulting series written onto a CD-ROM, along with an identifier for the feeder and the number of aggregated customers for both series. This CD-ROM will be sent to Component 23, a data store that provides a growing resource on joint probability distributions of aggregated demands on the same feeder.
- Continuing with our analysis of the journey of the Cranwood Feeder 7 case study through the tool, we saw in the Component 8 text above that the subset of customers for which data was requested in case are those 25 domestic customers connected to the red phase, and on the route from Way 7 to any of Nodes *F*, *G* or *H*. We imagine now that 8 of these customers have valid time series, so they are therefore aggregated and – since we imagine that the minimum number of customers for aggregation is 3, the resulting series is written onto the CD-ROM that will be supplied to Component 10, after the process has been conducted for every recognised feeder.
- Further, we imagine that 32 of the 116 domestic customers connected to the feeder have valid time series for the period. We imagine that these are randomly divided into groups of 9 and 23 customers, that the series in these groups are aggregated, and that both are written onto the CD-ROM that will be supplied to Component 23, after the process has been conducted for every recognised feeder.
- Non-domestic customer data does not need aggregating (with SLC10A only applicable to Domestic Premises). The first two digits of the MPAN indicate its profile class. Profile classes 01 and 02 relate to domestic premises, so consumption will be subject to aggregation rules. Profile classes 03 to 08 are non-domestic, so consumption will not be subject to aggregation rules.

Mode of Operation

- Periodic batch process.

Event Sequences and Triggers

- Triggered by the arrival at the Siemens EIP of a CD-ROM, from Component 8.

Data Inputs

- A table listing all the MPAN number for which data is requested, for each feeder represented by the tool, supplied on a CD-ROM with content written by Component 8.

Data Outputs

- CD-ROMs mainly comprising of aggregated 30-minute consumption time series covering 3-month periods, accompanied by the feeder identifier, a list of the MPANs that contributed to the series, and the total percentage of valid data in the individual series. The other content on the CD-ROM will be sets of histograms derived from such time series (where the number of MPANs involved in the aggregation is too small for the data to be shared), with the same labelling. This is sent to Component 10.
- CD-ROMs with the same type of time series data, but this time two series per feeder, labelled with the feeder identifier, and (only) the number of customers aggregated in each series.

Developments in Future Versions

- The basic operation is not foreseen to change significantly. It is hoped that a number of different data series, involving the demand of multiple customer groups for each feeder, will be exported by this component in the future.

4.11 Component 10: Smart Meter Data Integration Module

Functionality

- The purpose of this component is to accept and process the consumption time series data, and histogram data, provided to it via CD-ROM by Component 9.
- The first objective of this component is to decide whether the recently acquired smart meter data, for each modelled feeder, should replace the existing smart meter data for the same quarter and feeder combination, or should instead be discarded.
- In cases where the feeder has been added to the collection of modelled networks during the last quarter, or where there previously were no customers with smart meters (providing high quality data) within the selected customer subset, the decision is obviously to accept the new data. However, in the case where data already exists for a specific combination of quarter and feeder, the new time series will replace the existing one if there are more customers aggregated in the new series, so long as the % of data present in the new series is not substantially lower than for the existing series.
- Where the output from Siemens EIP for a particular feeder is a set of binned frequency counts, i.e. the numerical equivalent of a histogram, synthetical time series will be generated through random sampling. This is described in more detail below. For feeders where the number of customers with good quality smart meter data has increased since the corresponding quarter in the previous year, so that a direct time series could be extracted from Siemens EIP, whereas previously only synthetic series existed, the new 'real' series should replace the old, unless there is a very significant decrease in the % of valid data present.
- The exact decision rules here, essentially the relative weighting that should be given to increases in the customer numbers with smart meters, versus decreases in the % of data present, should be established as part of developing this component, guided by trends in both, as can be gleamed from existing data.
- The process of creating synthetic profiles will be based upon the assumption that demand values are uniformly distributed within the frequency count bins (i.e. 'numerical histogram' bins), so that the frequency counts can be converted into probability density functions for demand. These individual distributions, for each of the 48 half-hours of the day, can be combined with an assumption of Gaussian copulas - with constant coefficients estimated based on many examples from the CLNR data – to produce an approximate joint distribution for all time steps across the day. To produce e.g. a 90-

day winter season time series, the tool would simply sample from this joint distribution 90 times and join the day-long samples together.

- If the newly acquired time series satisfies the conditions for replacing the existing data, the next task for this component is to rearrange the new series in daily (48-step) segments, in such a way that it makes more sense for the recorded data to be combined with the 2.5-year-long historic consumption traces that constitute the CLNR domestic demand dataset initially used by this tool. The rearranging will be based on matching ranks, as is explained below.
- Consider the total demand time series obtained by aggregating every individual time series for all customers in the CLNR domestic demand dataset (after significant filtering down for high quality data), and a specific season form that series, e.g. the 1st of two instances of the Winter season. One might observe that, e.g. the daily maximum demand on the 2nd Tuesday in February was significantly *higher* than the 1st Tuesday in February. This is likely, in most cases, to be a consequence of a difference in weather between those Tuesdays, or perhaps because one followed a bank holiday.
- Consider also the total demand time series obtained by aggregating every time series provided by the most recent CD-ROM (that are already aggregated as small groups). Imagine that for this series, the opposite situation is true, i.e. the demand on the 2nd Tuesday in January was significantly *lower* than the 1st Tuesday. This would almost certainly reflect the fact that weather patterns were different on those different years. This difference should be accounted for and corrected to the greatest possible extent before Siemens EIP smart meter time series can be combined with the CLNR metered data. The approach to this problem to be adopted by this tool is to rearrange daily segments of the former series to match the ranking of daily maximum demand with that of the same season in the CLNR data. This applied to both the real and synthetic series Siemens EIP series.
- More specifically, a ranking series of daily demand peaks, for the aggregate demand of all high quality CLNR customers, should be produced for e.g. the first winter season in the CLNR measurement period. If the first day of that season had e.g. the 50th largest daily peak (of 90), while the 2nd day had the 63rd largest peak demand, the first two entries in the ranking series would be 50 and 63. A similar ranking series should be produced for the aggregate demand time series of all extracted Siemens EIP data for the most recent winter season. Imagine that for this series, the 50th largest daily peak demand occurred on the 76th day, while the 63rd largest peak demand occurred on the 18th day. That would imply that when constructing the rearranged, ranking-matched series of Siemens EIP data, the daily series segment that was previously 76th should now be first, the previous 18th day now the 2nd day, and so on. The same process must be repeated for all instances of seasons in the 2.5-year long CLNR measurement period. All of the individual series extracted from the Siemens EIP should be rearranged in this way.
- Following these rearrangements, for those feeders where it is decided that the newly acquired time series should replace the existing one, this tool must achieved this by removing and writing in Component 12.

Mode of Operation

- Periodic batch process, once every quarter.

Event Sequences and Triggers

- Manually triggered when the CD-ROM from Component 9 is received by the Data Administrator.

Data Inputs

- Time series and histograms labelled with feeder identifier, the MPANs of the customers involved, and the total % of valid data used in constructing the series. Supplied by CD-ROM from Component 9.
- MPANs and the total % of valid data used in constructing the series, for previously process and saved smart meter data, extracted from Component 12.
- Time series containing the demand ranking of each day in the historical seasons covered by the CLNR time series, as described above.

Data Outputs

- Time series labelled with feeder ID, number of aggregated customers and average % of customers contributing data, sent to Component 12.

Developments in Future Versions

- The basic functionality is unlikely to change, but the rules for which action to take might become more complex as the overall methodology becomes more sophisticated and robust.

4.12 Component 11: Annual Consumption Data Uploader

Functionality

- This component accepts ad-hoc uploads of annual energy consumption date, and associated MPANs. Being It filters out any clearly erroneous values, namely implausibly large positive consumption values, and replaces them with existing values for the same MPAN in Component 12. The replacements should definitely be made, as they represent more recent consumption patterns for the customers.

Mode of Operation

- One-off operation when triggered.

Event Sequences and Triggers

- Triggered by a manual upload of data through a simple interface.

Data Inputs

- Tables matching kWh values to MPANs, received via a physical medium such as a CD-ROM.

Data Outputs

- Tables matching kWh values to MPANs, sent as write commands to Component 12.

4.13 Component 12: Database for Annual Consumption and Smart Meter Data

Functionality

- Stores aggregated smart meter data provided by Component 10, annual energy consumption values provided by Component 11, and metered data for non-domestic customers provided by Component 13. Allows those components to read and write data.
- Stores the daily peak demand ranking time series derived from CLNR data, described in the Component 10 section.

- Also allows Component 15 to read data.

Mode of Operation

- Operates as a straightforward data store.

Event Sequences and Triggers

- No sequences or triggers, simply allows the read and write requests as described above.

Data Inputs

- Demand time series extracted from the Siemens EIP environment, and associated labelling, as provided by Component 10.
- The daily peak demand ranking time series for CLNR data, inputed before this component went live.
- Annual energy consumption values and associated MPANs, entered by Component 11.
- Any available metering data for non-domestic customers, provided by Component 13.

Data Outputs

- As above

Developments in Future Versions

- The intention is that network monitoring data will also be stored by this component, as a priority.

4.14 Component 13: Half-Hourly Supply Non-Smart Meter Consumption Uploader

Functionality

- Accepts ad-hoc uploads of any half-hourly supply non-smart meter consumption the DNO holds.
- The first two digits of the MPAN indicate its profile class. If the profile class is 00, this indicates that the premises has a half-hourly supply.
- This component filters out any clearly erroneous values, i.e. implausibly large consumption or export values – although these limits will be harder to define for non-domestic customers.
- When the data is first uploaded, this component will first check how many complete seasons it has (not necessarily with 100% data present), disregarding any series segments that do not span a complete season. It then reads stored data in Component 12 to check which of the seasons present in the newly acquired data already have data stored for that MPAN. If no data currently exists for the new data's combination(s) of season and MPAN number, the new data will be stored. If data does already exist for the combination(s) of season and MPAN, the new data will either be disregarded or over-write the existing data, depending on whether it has a higher percentage of valid data present compared to the existing data.
- However, before the data is saved, two things must happen. The first is that gaps in the data are filled-in: with interpolation in the case of very short gaps, or by using the average value for each 48 time-of-day within the sample for longer ones. The second stage of preparation is to reorganise daily segments of the series to achieve a matched rank sequence with the CLNR domestic demand series, almost exactly like in Component 10.
- Returning to our Cranwood Feeder 7 case study, we imagine that metered consumption data was available for each phase of the 3-phase customer, and therefore these three series were uploaded by this component to Component 12.

Mode of Operation

- Short periods of activity following a manual upload. At some point in the future some automated process involving software external to this tool may be developed, but activity would probably remain as short bursts.

Event Sequences and Triggers

- Triggered by a person uploading data through a simple interface, or possibly software external to this tool.

Data Inputs

- Half-hour resolution consumption time series, with associated MPAN. Uploaded manually, likely to arrive as large bundles.

Data Outputs

- Processed metered data sent as write commands to Component 12.

Developments in Future Versions

- In future versions, this component might be used for uploading and processing a wider range of 'other' data, such as techno-economic data such as from a Government census, and possibly Element Energy peak demand forecasts (or any eventual replacement of that tool).

4.15 Component 14: Store for Intermediate Model-Building Data

Functionality

- This component is a store for intermediate data relating to the models built by various components, as part of the 'journey' towards exceedance expectation values for specific nodes and line segments. It is used in relation to the network as currently built, in contrast to Component 21, which provides a very similar role for scenarios created by network designers.
- For each feeder modelled by the tool, this component stores:
    1. The full set of MPAN numbers, and their assumed phase connection (which is randomly allocated for some customers) – provided by Component 1. If it is known that any of them have EVs or heat pumps (HPs), this information will be recorded.
    2. The set of network variables (i.e. voltages at nodes and the thermal utilisation of cable sections, expressed as a 100% of their rated value) that has been deemed by Component 4 to require probabilistic modelling. These must be labelled in a way that corresponds to a location on the feeder, as represented by the network model for the feeder stored by Component 2.
    3. For each network variable identified by Component 4, a set of MPAN numbers identifying the customers who's demand must be aggregated to form the independent/regressor variable in the regression equation, where the network variable in question is the dependent/predicted variable. These MPAN numbers are calculated and placed in the store by Component 5.
    4. Again, for each network variable identified by Component 4, a set of parameters for the regression model relating the network variable in question to the aggregate demand of customers identified by Component 5. These model parameters are calculated and placed in the store by Component 6.

- This component allows components 4, 5, 6, 8, 15 and 17 to read the data stored in it, and the components 4, 5 and 6 to write data within it.

Mode of Operation

- Responds to read and write requests as they arrive, with effectively translates into series of batch operations, by a sequence of components, occurring at the end of each quarter.

Event Sequences and Triggers

- Responds to queries and write commands from components as stated above.

Data Inputs

- The set of input data is as described in the numbered points above.

Data Outputs

- Component 15 extracts the MPAN data placed by Component 1.
- Component 5 extracts the data on which network variables need to be modelled, placed by Component 4.
- Component 6, Component 8 and Component 15 extract the data identifying the MPANs associated with each modelled network variable, placed by Component 5.
- Component 17 extracts the regression model parameter data placed by Component 6.

Developments in Future Versions

- In future versions where more advanced modelling of generation is included, the details of any known generation associated with MPANs will be entered by Component 1. This is also true of any socio-economic data that may get introduced in future versions.
- If any future version represents the uncertainty associated with phase allocation by shuffling customers around, this would create significant challenges for the combined operation of this component and Component 8.

4.16 Component 15: Demand Variable-Set Sampler

Functionality

- This component constructs many samples of multi-year demand time series, to either represent the temporal dynamics of aggregated customer demands, or to represent the contribution of individual customer demands to typical annual peak demand conditions variables – depending on which component requested the sample. The former output will be delivered if triggered by Component 17, the latter if triggered by Component 4 or Component 6.
- More specifically, when this component is triggered by Component 4 or Component 6, they do so because they need a matrix of individual demand values, where the number of columns is equal to the number of MPANs associated with the feeder, and there are 1000 rows. Each row therefore represents the simultaneous demands of each customer connected to the feeder during some 30-minute time step. The matrix must be accompanied by a list indicating which MPAN number is associated with each column.
- As stated above the time steps we are interested in are those that represent peak demand conditions. This will be achieved by sampling 2.5-year-long series, extracted from Component 12 and Component 16, with one series representing each MPAN. The series will then be aggregated, before establishing the time and date at which the 10 largest aggregate demands occurred. This allows 10

rows of the matrix to be filled – one row for each of the 10 peak demand time steps, and one column for each time series (representing a specific MPAN). This process is repeated 100 times, drawing a different sample of time series (with replacement) each time.

- This process is illustrated by Figure 4.2, in the simplified hypothetical situation where there are only 3 connected customers, the time series only involve 20 time-steps, and we only extract the individual demand values corresponding to the top two highest aggregate demand values.
- When this component is triggered by Component 17, the sampled aggregate demand series will be used to fit parametric probability distributions for the network variables, after being passed through the appropriate network response model. In this case, instead of sampling a series for every customer connected to the feeder, each network variable identified by Component 4 will be modelled separately, and the number of sampled series will match the number of customers being aggregated in the regression equation for that network variable, as determined by Component 5.
- For the MVP, the presence of LCTs will be accounted for by adding generic demand / export profiles for each type of LCT to the CLNR TC1a demand time series.
- In all cases, i.e. when triggered by any of the components included above, the sampling process will proceed as follows. For each MPAN involved (both domestic and non-domestic), this component queries the metered data in Component 12 to see if a time series exists for the MPAN. If it does exist, the time series is extracted and placed in a temporary store inside this component.

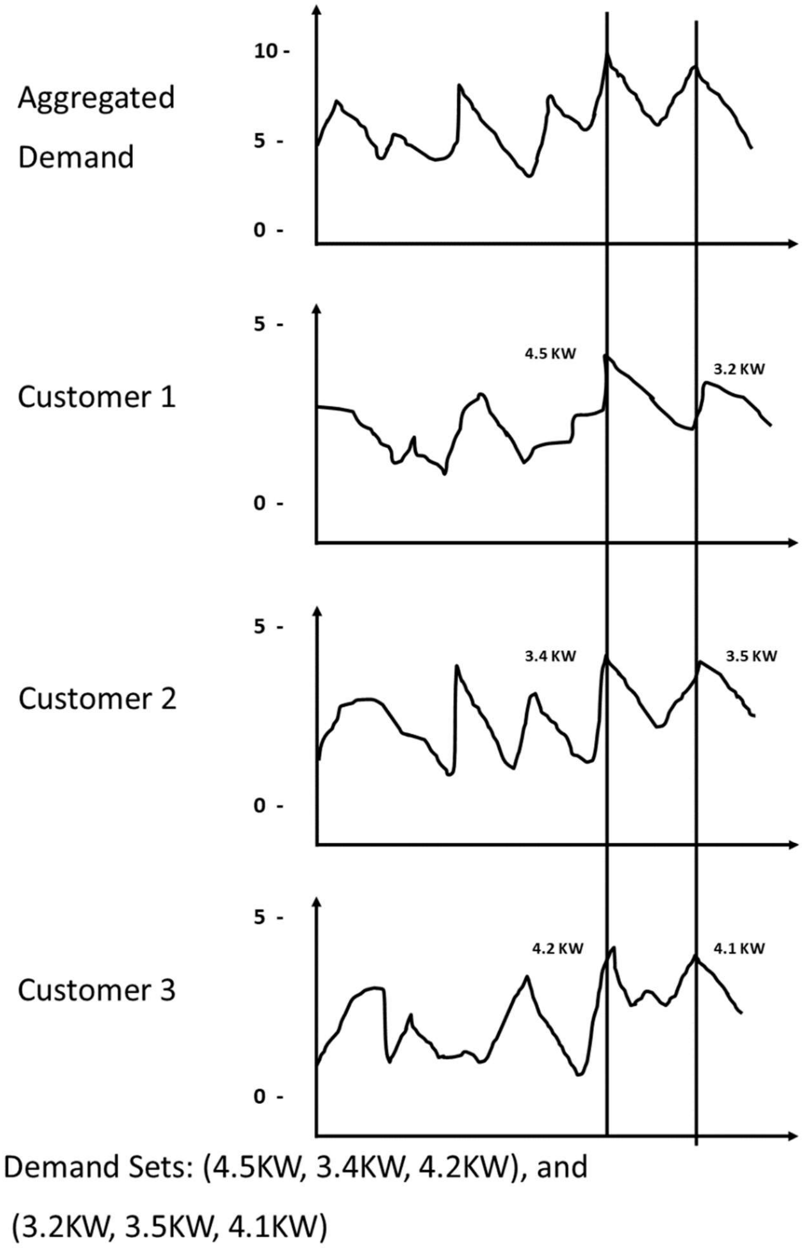

Figure 4.2. Demonstration of the principle used by Component 15 to populate demand matrices

- If the true monitored data for that MPAN does not exist, a CLNR domestic customer time series is randomly selected from the collection of these series stored in Component 16, as a stand-in. This process is repeated for each relevant MPAN, with each series stored separately in the temporary space, but also added together to create the final aggregated demand series.
- Another level of sophistication here is that use is made of the annual energy consumption data for the MPANs involved, which will be extracted from Component 12. That is, before the time series drawn from Component 16 are stored in the temporary space (and aggregated), they will be linearly rescaled so that they represent exactly the same annual energy consumption as the value extracted from Component 12 for that MPAN. Also, the CLNR time series will be sorted according to their calculated annual energy consumption, and put in 20 equal sized bins (of about 200 profiles each) on this basis. The profile for a particular MPAN will only be drawn from the bin with the annual

consumption range matching the MPAN's value, before 'fine tuning' the match through linear re-scaling. This ensures that e.g. a very heavy electricity user isn't represented with the demand pattern of a very light user, that's been re-scaled by a large factor.

- The treatment of PV generation, along with EV and HP demands, will be very simple for the MVP. In The only complication addressed is that they must match the times represented by the TC1a demand series.
- In the MVP, where metered data does not exist for non-domestic customers, they are simply represented as domestic customers, albeit probably scaled-up in comparison due to higher annual energy consumption values. This will change in future versions, where generic profiles for various types of non-domestic customers (e.g. schools, cafes, convenience stores) will be introduced.
- Returning to the Cranwood Feeder 7 case study, we recall that Component 8 requested smart meter consumption data from a group of 25 (out of a total of 116) domestic customers, and that Component 9 was able to provide high quality data for 8 of these. This means that when this component is instructed to work with our case study feeder by Component 4 or Component 6, it will draw the 8 correct domestic profiles, and 3 profiles (or perhaps one profile, assumed to be identical on each phase) for the non-domestic customer from Component 12 (rearranged and stretched to 2.5 years) and will draw 108 CLNR domestic time series form Component 16. After repeating this process 100 times, picking different CLNR profiles, the tool will be able to export a 116 x 1000 matrix of demand values. When triggered by Component 17 to work on Cranwood Feeder 7 and a specific network variable, this tool will again draw time series from Component 16, and possibly some from Component 12, depending on the network variable. The output will be a single aggregated demand series of length 2.5 years.
- The Cranwood Feeder 7 'base case', i.e. the network as it is currently believed to exist, is somewhat simple in that there are no assumed EVs or HPs. However, some other base case scenarios may assume that they are present, and certainly some of the hypothetical scenarios constructed by designers will contain these. When that is the case, EV and HP profiles stored in Component 16 will be randomly drawn and added to the domestic series.

Mode of Operation

- Can be batch operation every quarter, or ad-hoc when designers are testing hypothetical scenarios.

Event Sequences and Triggers

- Triggered to operate by Components 4, 6 or 17. Likely to be triggered several times sequentially for the first two, and 100 times for Component 17.

Data Inputs

- Extracts time series from Component 12 and Component 16.
- Extracts annual energy consumption values from Component 12.
- Extracts either the full set MPAN numbers, or those associated with a specific network variable, from Component 14, depending on which Component triggered it.

Data Outputs

- Demand time series (kW), either individual customers or aggregated or a matrix of peak values.

Developments in Future Versions

- A much more rigorous representation of power export time series, mainly PV, along with generic profiles for various types of non-domestic customer commonly connecting at the LV feeder level.

4.17 Component 16: CLNR Consumption Time Series Store

Functionality

- A database storing all of the data needed by Component 15 to produce the demand time series samples.
- It also stores some generic, but empirically derived EV and HP demand time series, and PV export time series.
- It will be populated manually, once with all the relevant CLNR project datasets – TC1a, TC3 and TC6, with TC1a (domestic customer profiles without LCT) by far the most heavily used. Before the TC3 and TC6 series (HPs and EVs, respectively) are uploaded, they must be rearranged and 'stretched' to 2.5 years using the ranking series matching approach described in the text for Component 10.
- As described in the text for Component 15, the TC1a series are placed in 20 bins of equal size, defined by intervals of the annual energy consumption that the series represent.

Mode of Operation

- A database responding to read requests. Effectively means it will initially be in batch operation, and later responding as part of a chain of events when network designers are using the tool.

Event Sequences and Triggers

- Responds to queries from Component 15.

Data Inputs

- Demand time series (kW) – the bulk being CLNR TC1a consumption data, but also some generic LCT profiles

Data Outputs

- Demand time series (kW)

Developments in Future Versions

- Future versions will probably store customer-specific profiles for embedded generation, and several types of non-domestic customers, but when more data becomes available, they will be integrated into the modelling in the same way as domestic smart meter data.

4.18 Component 17: Probability Distribution Fitter & Exceedance Expectation Calculator

Functionality

- To calculate exceedance expectation values for each modelled network variable (voltage or thermal utilisation) of interest on every main feeder, given the demand-related data stored in other components.
- The stages of doing this, for each modelled network variable, on each feeder, is as follows:
    1. Extract the relevant modelling data from Component 14, including regression model coefficients;
    2. Instruct Component 15 to produce a sample of the required aggregate demand time series;
    3. Use the regression model coefficients to transform the aggregated demand series into a time series for the network variable.

4. Fit a series of parametric probability distributions to the time series – distinct distributions for each time of day and each season, which will be a combination of Gamma and 3-parameter Weibull distributions. This model fitting process is presented in detail in the Project's report on Smart Meter Data Analytics[4], and is too complex to be presented here. We assume here that distinguishing between weekend and weekdays is not necessary for the first version of the tool.
5. The fact that Gamma distributions only support non-negative variables might initially seem like a problem for this component, given that net demands are negative when generation is greater than demand, and that the 2-parameter versions of the distributions do not support negative numbers. Fortunately, this can be resolved by an additional function within the component that shifts the net demand series so as to always be non-negative before the distribution-fitting process, followed by a reversal of this shift.
6. Assuming the thresholds of interest are a thermal utilisation of 100%, and a voltage drop at 100% of the allowed limit (the first version will only consider low voltages, as generation is not being represented) – the probabilities of these levels being exceeded must be calculated separately for each hour of day and season. These probabilities are given by 1 – F(x), where the F(x) are *cumulative probability distribution functions*, which are known analytical expressions for parametric distributions. The thresholds can, if desired, be changed to e.g. 110% of the allowed limits, if 100% is deemed too 'mild' to represent a serious problem, which might be reasonable for short thermal overloads. We refer to this pair of values (one for thermal, the other for voltage) the exceedance thresholds.
7. Once the set of probabilities are calculated, they are multiplied by the number of such periods per year – e.g. the periods 5:00-5:30pm in winter (December to February) occur 90.25 times a year. The results of this multiplication, conducted for each combination of time of day and season, are added together to obtain the *Exceedance Expectation* for the network variable being modelled (for a given exceedance threshold).
8. Steps 3 – 7 are repeated 100 times, to obtain a set of 100 exceedance expectation values. The mean of this set is taken as the final *Exceedance Expectation* value, and may be inverted to provide the *expected number of years between exceedance events*, e.g. concluding that "a voltage violation at this node is a 1-in-8.5-year event.

- The process outlined above is repeated for every network variable identified as needing to be modelled, and when all variables on a feeder have been calculated, the results are saved in Component 18. It may also be worthwhile storing a binary warning variable in each case, indicating whether the exceedance expectation is above some threshold value set by the DNO (e.g. 1-in-10-years). These values will be presented to network designers as part of a visual interface put together by Component 19.
- When working on a network as it is currently believed to be, i.e. base case scenarios, results are stored in Component 18, whereas they will be stored in Component 21 when calculating results for hypothetical scenarios.
- Returning to our case study of Cranwood Feeder 7, this component would run calculations on the three identified variables: the thermal utilisation of all three phases of the first cable section connecting to the way, the voltage of the red and yellow phases at node *E* and the red phase voltage at node *G*. We imagine that the expected frequency of exceedance for the voltages at node *E* is 1-in-12 years and 1-in-16 years, respectively, while the red phase voltage at node *G* has an expected

[4] Available at: https://www.northernpowergrid.com/asset/0/document/4803.pdf

frequency of 1-in-13 years. We further imagine that NPg's policy for acceptability for voltage infringements is no more frequently than 1-in-10 years, so none of voltage results demand network reinforcement. We imagine that the thermal utilisation results for the red, blue and yellow phases were frequencies of occurrence of 1-in-9.5-years, 1-in-11-years and 1-in-10.5 years, respectively. If NPg's policy is no more frequently than 1-in-10 years, then the red phase result would mean that reinforcement is necessary. However, NPg's policy could be slightly more nuanced, for example allowing the expected frequency for 100% utilisation to go up to 1-in-9 years, as long as the expected frequency for a utilisation of 105% is less frequent than 1-in-10 years. In that case, we imagine that the network planner re-runs calculations for that particular variable at 105%, and that the result turns out to be 1-in-11.5 years – and so the network needs no reinforcement.

Mode of Operation

- Initial batch operation, working through each variable for each main feeder for all new LV networks represented by the tool. Triggered to do this by Component 6.
- The other mode of operation is to respond to network designers using the system, i.e. responding to triggers from Component 19.

Event Sequences and Triggers

- Conducts batch operation on all new or modified feeders, once all of the required data is in place. Being triggered to enter batch operation when Component 6 has completed its own batch operation is a way of ensuring this. It must be noted, however that a batch operation on a number of recently added feeders could potentially be triggered e.g. one week before the smart meter data is updated (via CD-ROM), which would render the data out of date very quickly. We believe that this is an acceptable limitation due to the complexity of any alternative.
  Responds to triggers from Component 19, as part of a carefully coordinated response from several components when a designer wishes to test a hypothetical scenario.

Data Inputs

- Modelling data from Component 14, aggregate demand time series from Component 15.

Data Outputs

- Exceedance expectation values, and possibly a yes/no indicator for very frequent exceedances, sent to Component 17.
- Probability distribution parameters and covariance matrices, sent to Component 18.

Developments in Future Versions

- This component, along with Component 15, are expected to change very radically in the future, with the full adoption of a Bayesian methodology, which is the only statistical approach that can robustly handle the subtleties and complexities faced by this tool. However, the adoption of a fully Bayesian approach is not currently possible, due to very limited data available, particularly in relation to the joint distribution of two aggregated demand variables that are actually on the same feeder – the sampling approach is forced to make the strong assumption that the statistical relationship between the demand patterns of domestic customers is not a function of their geographical distance. A detailed presentation of TNEI's proposed Bayesian approach, along with the simpler, sampling-based approach specified here is provided in the Project's Novel Analysis Techniques Report[5].

4.19 Component 18: Exceedance Expectations Store

Functionality

- Stores exceedance expectation values for every modelled network variable (utilisation and voltage), and makes them available to network designers Component 19. These values are for the as-built network only, with results for hypothetical scenarios tested by designers stored in Component 21.

Mode of Operation

Operates as a database, accepting queries from Component 17 and Component 19, and write requests from Component 17.

Event Sequences and Triggers

- Responds to queries and writing commands, thus effectively initial batch operation, followed by more ad hoc use, when designers are examining the base case scenario for a particular feeder.

Data Inputs

- Exceedance expectation values, plus flag for extreme values

Data Outputs

- As above

Developments in Future Versions

- This component would expand significantly to store many more parameters and hyper-parameters, including for covariance matrices, following a transition to a fully Bayesian approach.

4.20 Component 19: Smart LV Design Web Application

Functionality

- The purpose of this component is to coordinate and execute the commands of the tool's users, so that they can view the extent of any overloads in any chosen feeder for the network as it is currently believed to be (base case scenarios). It also enables them to create, test and save new hypothetical or forecasted scenarios.
- Interactions occur via a web browser, Component 20, but that Component is essentially a delivery mechanism for the outputs of this component, and a means of transmitting mouse clicks and entered text into commands for this component.
- It will reside on NPg's Corporate IT network and will interface with this tool's dedicated Enterprise Server (where most components reside) via an Enterprise Bus[6].
- The outputs passed to Component 20 will be the data needed to construct a schematic diagram of a feeder, annotated with 1/ (exceedance expectation) values for each network variable identified as

[5] Available at: https://www.northernpowergrid.com/asset/0/document/4918.pdf

[6] The Enterprise Bus is a middleware product which will be used as the message communication mechanism between the Corporate IT Network and the Smart Grid Enterprise Server.

requiring modelling, and on a geographically accurate map background. This component uses GIS data from Component 22 to create that map background.

- The user commands passed to this component by users can be classified into a number of types. These command types are listed below, followed by the actions required from other components in order to execute them:
    1. Display a feeder or entire LV network of the user's choice, including pre-calculated base-case exceedance expectation values. This should include – possibly as popups that following a mouse-click – all presumed customer demand details for a feeder, e.g. the presumed location of LCTs, and the nature and location of non-domestic, non-LCT demand.
    2. Facilitate browsing through the most overloaded feeders – defined as those with network variables with the highest exceedance expectations, for a threshold level of 100% - in decreasing order of exceedance expectations. Again, these are for the base-case scenario and should therefore be pre-calculated and stored (with the exception of the period immediately after new data has been uploaded into the system, triggering batch operation of many components).
    3. The same as (1) above, except for a hypothetical scenario previously created and saved by a tool user.
    4. Display all details of network components, e.g. power rating, material and cross-sectional area for a cable – for a selected subset of network components.
    5. Re-calculate exceedance expectations for a feeder of the user's choice, for a different threshold level than the last time calculations were ran for that feeder, but without changing demand patterns or the physical network. Save as a new scenario if desired.
    6. Re-calculate exceedance expectations for a feeder of the user's choice, where there are changes to demand patterns compared to the last calculations for that feeder, e.g. EVs added, but where the network is unchanged. The exceedance threshold level may or may not be the same as for the last calculation. Save as a new scenario if desired.
    7. Re-calculate exceedance expectations for a feeder of the user's choice, where there are changes to the physical network compared to the last calculations for that feeder, e.g. cable section reinforced, but where the demand patterns are unchanged. The threshold level may or may not be the same as for the last calculation. Save as a new scenario if desired.
    8. Re-calculate exceedance expectations for a feeder of the user's choice, where there are changes to the physical network compared to the last calculations for that feeder, *and also* changes to the demand patterns. The threshold level may or may not be the same as for the last calculation.

- Following a command of type (1) above, this component must: retrieve the network and customer data for the chosen feeder from Component 2, the calculated exceedance expectation values from Component 18, and the relevant GIS data from Component 22. It will then integrate and process the data to construct an annotated diagram, and send it to Component 20.

- Following a command of type (2) above, this component must: begin by querying Component 18 to find the highest exceedance expectation value. It will then construct and export the annotated diagram in the same way as for type (1), and noting the feeder ID. If the user wishes to see the 'next worse' overloaded feeder this process will be repeated, but where the recently constructed feeder IDs are excluded from the search for the highest exceedance expectation value.

- Following a command of type (3) above, this component must: retrieve all of the data required to build the model from Component 21, then the necessary GIS data from Component 22. It will then integrate and process the data to construct an annotated diagram, and send it to Component 20.
- Following a command of type (4) above, this component must: do exactly the same as for command type (1), but extract more data from Component 2, namely the conductor specifications, and use this to construct a more detailed diagram, sent to Component 20.
- Following a command of type (5) above, this component must: provide Component 17 with all of the relevant information for the feeder (and scenario) and instruct it to run calculations with the chosen pair of exceedance thresholds. If starting with a non-base case scenario, the data required by Component 17 will be drawn from Component 21. Regardless of the starting scenario, the destination for saving all the data required by Component 17 is Component 21. The data will again be integrated with GIS data from Component 22 and used to construct an annotated diagram sent to Component 20.
- Following a command of type (6) , (7) or (8) above, this component must: start 'from scratch' in calculating the relevant exceedance expectations, including using Component 4 to check whether any additional components might now satisfy the conditions for requiring probabilistic modelling; also using Component 6 to check if the coefficient of the regression equation has changed. It will be possible to omit some detailed aspects of the calculations, depending on which of the commands was issued. If the scenario being modified was base case, the initial network and basic customer data will be retrieved from Component 2, but as the analysis progresses through the chain of components, all data storing – including intermediate modelling data – will be stored in Component 21. If the scenario being modified was not a base case, then all data retrieval and storage, as the analysis progresses, will involve only Component 21 only. As for the other commands, data will be integrated and processed so that an annotated diagram can be passed to Component 20.

Mode of Operation

- Operates only when a network designer is using the tool, as indicated by triggers from Component 20. Causes a number of other components to operate in reactive mode.

Event Sequences and Triggers

- Receives triggers, followed by data, from Component 20. Triggers several components, depending on the command received, as outlined above.

Data Inputs

- Requests for data to be retrieved or re-calculated.
- Details of changes the user wishes to test, which can take the form of any combination of (i) changes to the network, e.g. reinforced line segment; (ii) changes to demand patterns, such as adding EVs at specific nodes; (iii) a change to the exceedance threshold.

Data Outputs

- Data required to form a schematic diagram of the feeder.
- Values of 1/ (exceedance expectation), annotating the schematic diagram.
- Earth loop impedances (not part of the main flow of calulations)
- Possibly conductor specifications, for annotating the schematic diagram.

Developments in Future Versions

- The basic functions of this component will remain unchanged in future versions of the tool. However, many details will change significantly with the adoption of a Bayesian approach.
- Future versions of the tool should ideally be extended with the option of adding network management tools that are smarter than network reinforcement, e.g. demand response, automatic voltage control algorithms.

4.21 Component 20: Network Designer User Interface (Web Browser)

Functionality Summary

- This component is a web browser that provides network designers, or any other potential user of the tool, with a means of accessing and interacting with the tool components inside the tool's dedicated Enterprise Server, from their personal computer terminals, within the DNO's wider corporate IT network.
- All interactions are mediated by Component 19, and were therefore described in that section – the purpose of this component is simply to render images and allow commands in the form of entered text and mouse clicks to be conveyed to Component 19.

Mode of Operation

- A responsive interface for the tool's users.

Event Sequences and Triggers

- Takes input from users and presents them with a visual representation of feeders, along with the results of calculations.

Data Inputs

- Mouse clicks and drags, entered text from users.
- Data received from Component 19, as a response to commands given to it.

Data Outputs

- Schematic diagrams of feeders, annotated with values, presented to users. By default, the annotated values are 1/ (exceedance expectation), but also optionally network component details and customer demand types.
- Graphical interfaces to allow users to change the customer demands or network component specifications
- Commands for data retrieval or re-calculation issued to Component 19.

Developments in Future Versions

- Any changes to the nature of the tool's interface and functionality are really changes to Component 19, rather than this component.

4.22 Component 21: Smart LV Design Database

Functionality

- This component is required so that network designers can store hypothetical scenarios – involving changes to existing feeders and the customers connected to them – so that they persist between sessions. These scenarios might be made to differ from the base case scenario by adding new circuits, upgrading circuits, or adding new demand (and, in future versions, generation). It shall be possible for the user to share these study cases with other users in their workgroup.

- This component is a data store, storing a variety of variables that are spread across several data store components in the case of base-case scenarios. The data stored must be sufficient to enable other components to function, so that previously saved hypothetical scenarios can be shown, including all modelling results, to designers – who may choose to make changes to the scenario before running all calculations on the new version. The database will therefore inevitably have quite a large schema with many of the data structures copied from the other data stores.
- The details of the data it accepts from and provides to various components are presented in full detail in the sections of the Components concerned, and are not reproduced here, in the interests of brevity.

Mode of Operation

- Receives queries from multiple components whenever a network designer, or any other user, wishes to create and test their own scenario, or examine and edit scenarios previously saved by themselves, or another designer within their working group.

Event Sequences and Triggers

- Allows components 5, 6, 7, 17 and 19 to make queries and allows components 4, 5, 6, and 19 to enter new records or overwrite existing records.

Data Inputs

- Multiple datasets stored in this component are extracted by components 5, 6, 7, 17 and 19, with the details provided in full in those components' sections.

Data Outputs

- Data is written into many tabled stored in this component by components 4, 5, 6, and 19, with the details provided in full in those components' sections.

4.23 Component 22: Ordnance Survey Master Map

Functionality

- This is essentially a store of GIS data that enables Component 19 to present users with graphical representations of the LV feeders, or entire LV networks, that are spatially accurate and placed in a geographic context.
- It will be pre-loaded with data when the tool goes live, and allows Component 19 to access that data when needed.
- Should allow manual updates, e.g. to reflect a new housing estate being built.

Mode of Operation

- Operates whenever a user wishes to view or edit a feeder or entire network.

Event Sequences and Triggers

- Provides Component 19 with GIS data when required.

Data Inputs

- GIS data uploaded once, before the tool goes live, with occasional manual updates.

Data Outputs

- GIS data.

4.24 Component 23: Joint and Marginal Probability Distribution Learning Database

Functionality

- This component is quite separate from all others, and does not contribute to the functioning of the tool. Rather, the purpose is to capture data that's necessary for future development stages of the tool, in particular the transition to a fully Bayesian approach.
- The presence of PV generation, and possibly other distributed generation types, adds some complexity to the data captured by this component, although the MVP will keep matters as simple as possible.
- Storing pairs of aggregated consumption time series, for groups of customers on the same feeders, allows future analysis of joint probability distributions of demand for such paired groups – i.e. accurately capturing the nuances of the statistical relationship between the demands of groups of customers on the same feeder, as well as their individual probability distributions. This is necessary on order to obtain Bayesian prior parameter distributions, essential to adopting a Bayesian approach – the benefits of which are discussed in detail in our report on Novel Analysis Techniques for this project.

Mode of Operation

- Operates as a database, with new records being created every quarter, as a batch process.

Event Sequences and Triggers

- Triggered manually to read the contents of CD-ROMs and add them as new records in the database, when the CD-ROM is available to the data administrator

Data Inputs

- Contents of CD-ROMs, comprising of pairs of aggregated 30-minute consumption time series covering 3-month periods, accompanied by the feeder identifier, and the number of customers aggregated in each series.

Data Outputs

- The stored data will be extracted and analysed as some point in the future, as an essential part of making a transition of the tool from sampling-based analysis to more accurate and efficient analysis based on Bayesian updating.

Developments in Future Versions

- Even after a transition to a fully Bayesian approach, it will be worthwhile to collect the same data, and indeed to capture the statistical relationships between multiple demand variables on the same feeder in more sophisticated ways, particularly during extremes, both high and low, of overall demand on the feeder.

5 Assessing Alternative Scenarios

This section consolidates understanding of how network designers would interact with the tool, by presenting a number of scenarios. These are examples of modifications a designer could make to the Cranwood Feeder 7 base-case scenario, and we present here an imagined narrative of the tool's analytical results for them.

5.1 Scenario 2

For this scenario, we imagine that a designer wants to model the situation if 15 single-phase, 16 amp, on-street EV chargers connected to the network in an area known as Skip Worth Close (labelled in Figure 5.1). The designer therefore adds these to the base case demand model, saving the result as a new scenario. All the data needed to recreate and model this scenario will be stored in Component 21. The analysis would be run through the full sequence of analytical components with the additional demand.

In fact, due to the fact that the MVP will not provide results that are probabilistically balanced over all possible states of phase imbalance, we imagine that the network designer produces two scenarios with contrasting states of phase imbalance. That is the designer produces Scenario 2A, in which the EVs are added completely evenly across the phases, and scenario 2B, which is relatively extreme example of phase imbalance, with the imbalance of the EVs worsening the existing imbalance. The details of how the new demand is connected to the phases is shown in Figures 5.1 and 5.2 below.

We imagine that for both versions of scenario, Component 4 did not identify any additional network variables as requiring probabilistic modelling. Thus, the only change in the process of modelling expected rates of occurrence for exceedance, compared to the base-case scenario, is that Component 15 must add a sampled EV demand time series to the sampled domestic demand time series for a subset of customers. Also, average annual energy consumption values for 16 amp EV chargers would have to be added to the specific annual consumption values stored for the relevant MPAN values. Since the EVs are not connected to the part of the network for which smart meter data from the Siemens EIP system is requested, it is valid to re-use the smart meter without adjustment.

We imagine (using artificially convenient numbers that the results of the tool's calculations for Scenario 2A is that the expected rates of exceedance for the 3 voltage variables have increased as follows: 1-in-11 years and 1-in-15 years for the red and yellow phases, respectively, at node *E*, and 1-in-13 years (unchanged from base case) for the red phase at node *G*. For the three utilisation variables, the expected exceedance rates are 1-in-8.5 years, 1-in-10 years and 1-in-9.5 years, for a threshold of 100% utilisation. However, having increased the threshold to 105% utilisation, the expected frequencies were 1-in-10.5 years, 1-in-14 years and 1-in-12.5 years – therefore the designer concludes that no reinforcement would be necessary for this scenario.

However, for the more unbalanced arrangement in Scenario 2B, we imagine that the following results were found. Red and yellow phases voltages at node *E*: 1-in-9 years and 1-in-10.5 years, respectively; red phase voltage at node *G* unchanged at 1-in-13 years; the red, blue and yellow utilisation variables at 100% threshold: 1-in-7 years, 1-in-10.5 years and 1-in-8 years, respectively; for 105% utilisation threshold, the red, blue and yellow utilisation variables: 1-in-9 years, 1-in-11.5 years and 1-in-10 years. Given that the thermal utilisation results are very close to breeching policy for the balanced scenario, and breeches it for the unbalanced scenario, the designer would conclude that the arrival of the 15 EV chargers would necessitate reinforcement of the 0.1 SQ IN Al 5 Core cable feeding Skipworth Close.

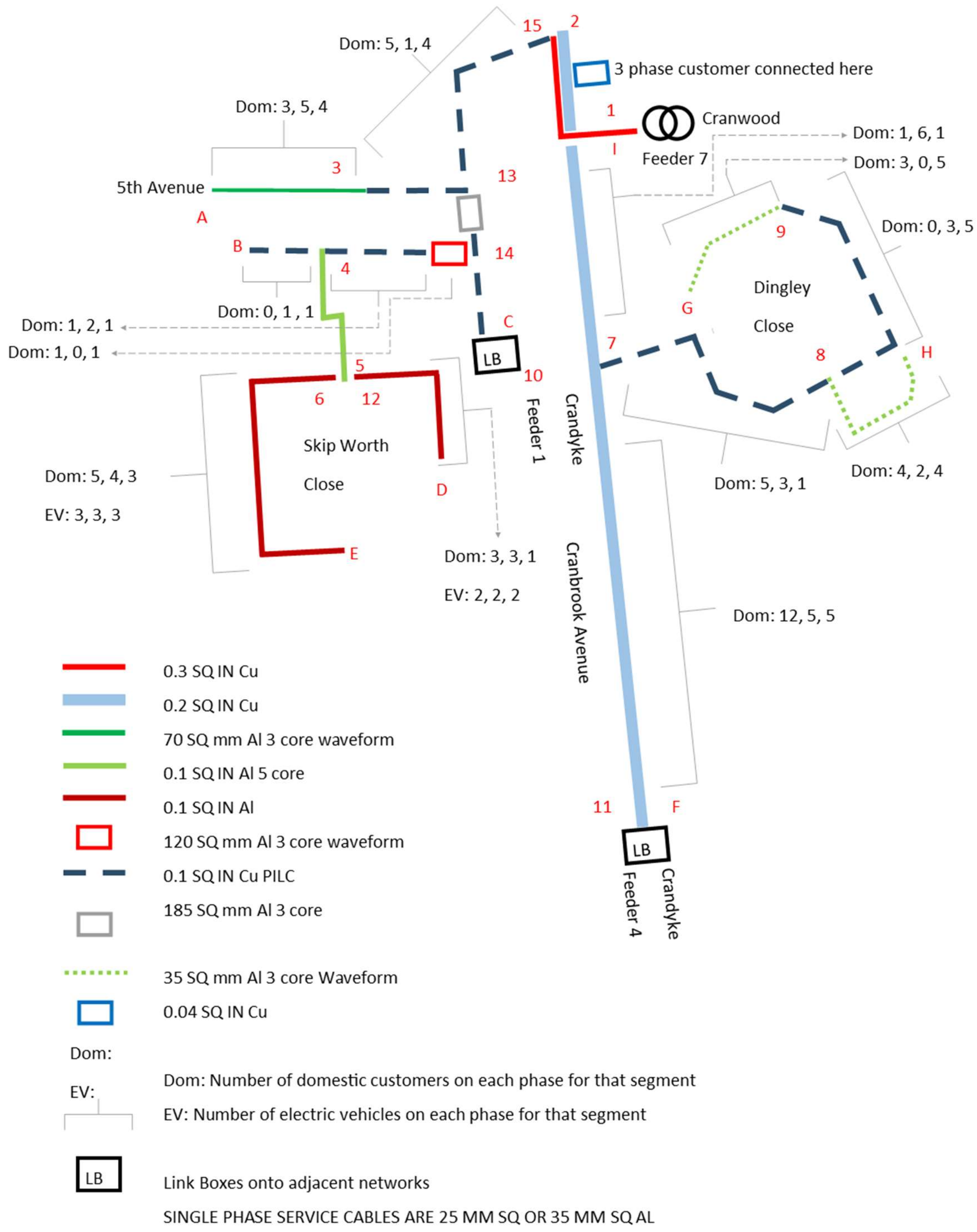

Figure 5.1. Demand Scenario 2.1 for Cranwood Feeder 7 – EV connections at Skip Worth Close are evenly balanced across the phases

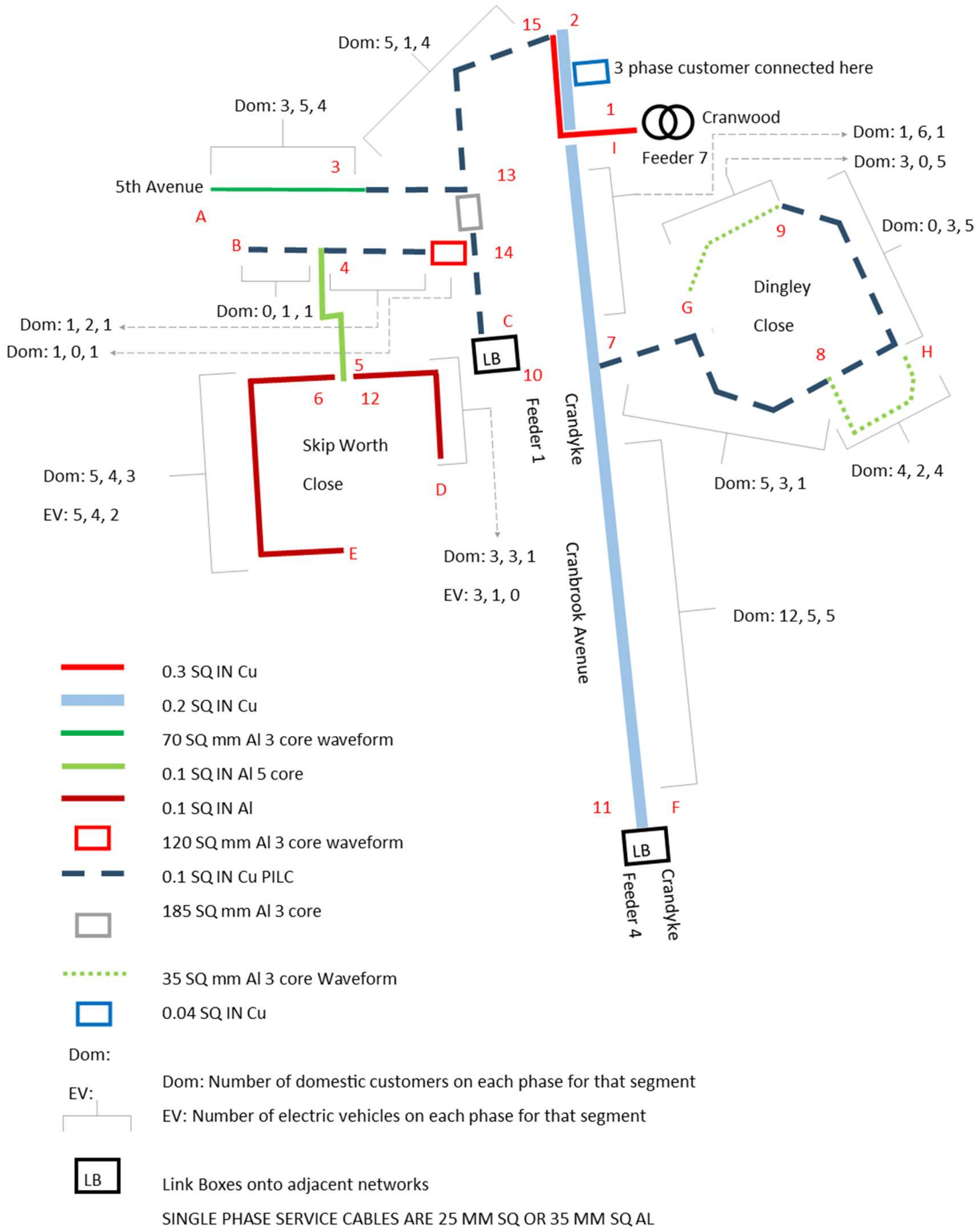

Figure 5.2: Demand Scenario 2.2 for Cranwood Feeder 7 - EV connections at Skip Worth Close are unevenly balanced across the phases

5.2 Scenario 3

Scenario 3 represents the situation where the additional EVs modelled in Scenario 2 have materialised on the network, along with an infill development of six new houses. The houses are added to an area called 5th Avenue, as labelled on Figure 5.3, and the network is extended from node *A* in order to serve the new customers. Specifically, the new cable is 100m of 300mm^2 Al Waveform cable. So, in this case, we are making simultaneous changes to the network and to demand patterns. For simplicity of presentation, we ignore phase connection uncertainty for now, and imagine that it is exactly the balanced arrangement shown in Figure 5.3 below.

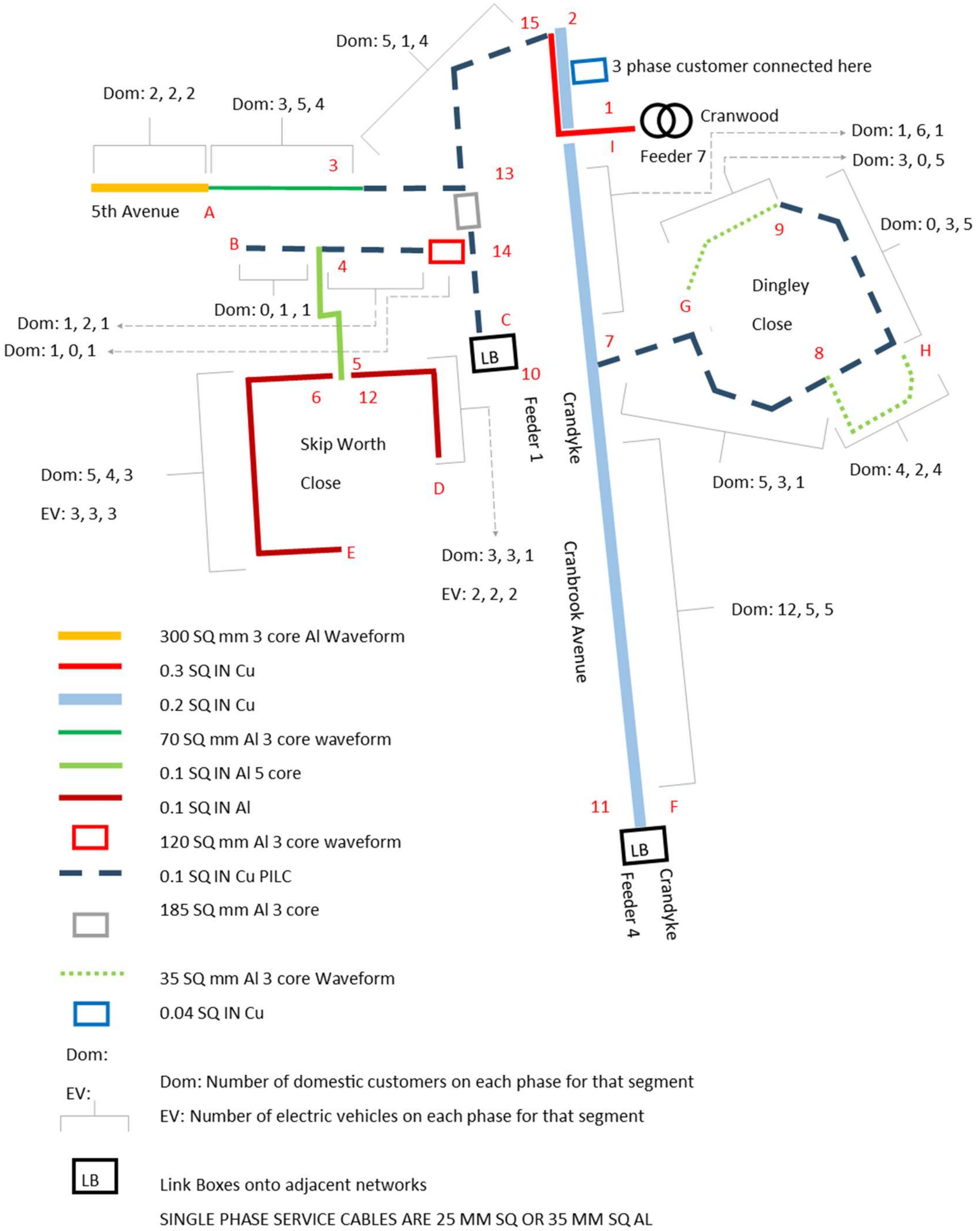

Figure 5.3: Demand Scenario 3 for Cranwood Feeder 7.

When the designer runs the calculations for this scenario, we imagine that Component 4 identifies the red phase voltage at the downstream end of the new cable as an additional network variable requiring probabilistic modelling. The components then work together to perform the probabilistic analysis on this extended set of variables. As before, the location of the changes compared to the base case mean that the base-case Siemens EIP smart meter data remains valid. As for Scenario 2, Component 15 must add EV samples to the domestic demand samples for the relevant MPAN numbers, and the relevant annual energy consumption values will be increased by the average value for 15 amp EV chargers. Additionally, Component 15 will use domestic CLNR time series samples to represent the demand of the new houses, and they will be assumed to have average annual energy consumption values for domestic properties.

We imagine that, when rounded to the nearest half year, the expected rates of exceedance for each previously existing network variable is unchanged compared to Scenario 2A, while the new voltage variable has an expected rate of 1-in-14.5 years. Therefore, the designer concludes that no network reinforcements are necessary for this scenario.

6 Roadmap to Implementation

Some ideas for ensuring the successful, gradual implementation of the tool – initially by NPg, and then by other DNOs, are the following:

- Follow the principles of Agile development and incremental roll out.
- Use Cranwood, Crandyke and Sinderby as initial test networks, and then increasing numbers of networks within the Yorkshire licence area.
- Make the software for the core functionality open source. Specific adaptors for different DNO environments. e.g. adapter to Spatial would be NPg-specific.
- While the purpose of the tool is to model LV networks much more accurately, and with the ability to incorporate multiple sources of data, ensure this does not come at the expense of unfeasible calculation times. Acknowledge that satisfying the requirements for complexity, accuracy and speed might necessitate greater computational resource that presently used, but contextualise with the scale of change and uncertainty facing networks.
- Be proactive in keeping other DNOs engaged and informed of progress, emphasising the very considerable novelty involved.