Investigating the sensitivity of network loss estimation to data accuracy and fidelity issues

Data Analysis

Ilias Sarantakos
David Greenwood
Peter Davison
Charalampos Patsios

Table of Contents

List of Abbreviations

DSO	Distribution System Operator
CLNR	Customer-Led Network Revolution
SCADA	Supervisory Control and Data Acquisition
WP	Work Package
ACE	Activating Customer Engagement
LCT	Low Carbon Technology
SME	Small Medium Enterprise
LV	Low Voltage
HV	High Voltage
EHV	Extra High Voltage
I&C	Industrial and Commercial
OLTC	On Load Tap Changer
CREST	Centre for Renewable Energy Systems Technology
RTTR	Real-Time Thermal Ratings
UGC	Underground Cable
OHL	Overhead Line
MTTF	Mean Time to Failure
MTTR	Mean Time to Repair

1 Introduction

1.1 Project Context

Enhanced Understanding of Network Losses is a project which seeks to enable DNOs to better understand, and my decisions pertaining to, the unavoidable losses which take place in their networks. The project is broken down into five work packages:

- WP1: A literature survey which outlined the state of the art in loss estimation, the methods and tools for managing and reducing losses, and the key issues which need to be addressed by future research, including this project.
- WP2: A data acquisition and analysis exercise, primarily comprising sensitivity analyses, which seeks to identify the key drivers of losses, and the network and measurement parameters which dictate how accurately losses can be estimated.
- WP3: A modelling exercise which will enable a limited number of representative models and methods to provide learning which can benefit the majority of distribution networks in Great Britain
- WP4: Using the methods from WP3 and the data and learning from WP2, to investigate future scenarios in which changing demand and new technologies are introduced into the network, and the impact this has on network losses. This will be carried out in a four level approach: Level 1 will investigate demand growth; Level 2 will look at the uptake of low carbon technologies, which will fundamentally alter the demand patterns present in the system now; Level 3 will look at smart grid actions taken without considering the impact of losses; and Level 4 will investigate how smart grid actions can be altered, by including losses in the decision making process.
- WP5: To propose policy and regulatory measures to help incorporate losses into decision making, and particularly how losses should be viewed in a system with electricity whose cost and carbon intensity varies with time and location.

An overview of the project, the specific work which will take place at each stage of the project, and how these are interlinked, is shown in Figure 1. This report describe the activities completed in WP2 and the learning which will be taken forward into the later stages of the project.

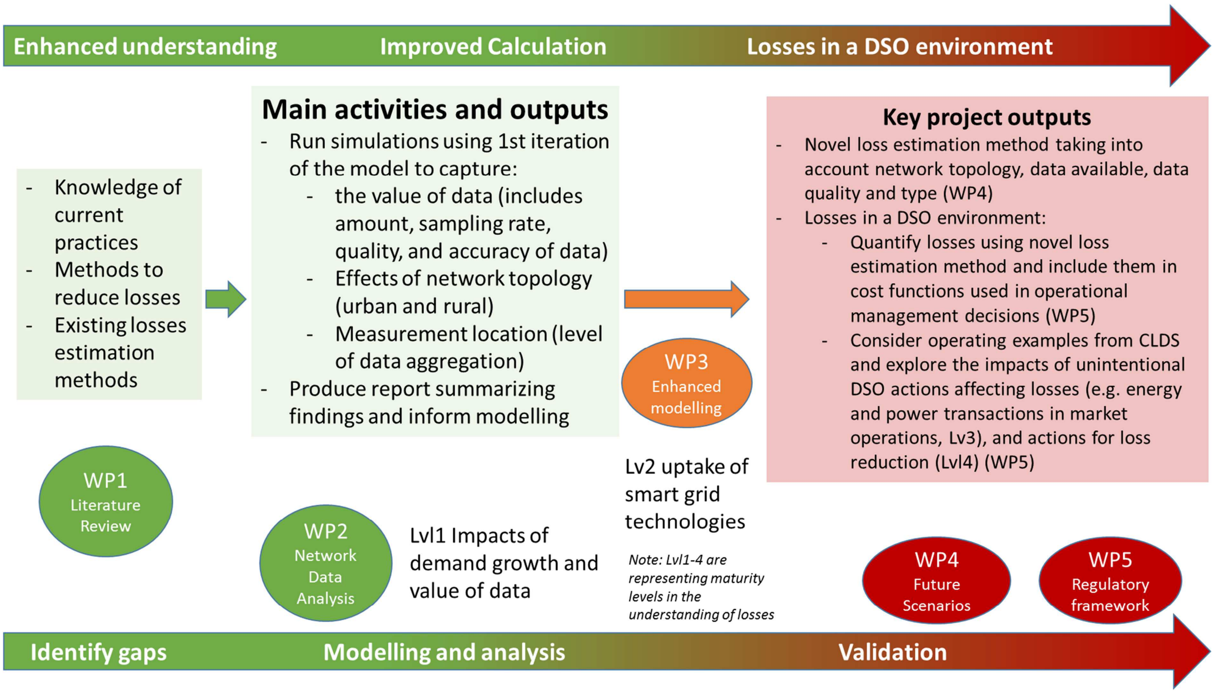

Figure 1: An overview of the project work packages and objectives

In a previous report, WP1 reviewed the state of the art in loss estimation and loss reduction in both academic literature and industrial practice, discussing methods for resolving the variability of demand; estimating the impact of low-carbon technologies, including harmonic currents; the impact of load imbalance; methods for estimating the impedance of Low Voltage networks; and assessing the impact of measurement error and data granularity on the accuracy of loss estimation. The key findings have been summarised in four categories i.e. the impact of present and future network scenarios including the variability of demand and generation, the impact of Smart and non-Smart Technology, the impact of measurement errors, and the impact of measuring at multiple aggregation levels. In a subsequent report and paper, we revealed the effects of demand growth on losses and their estimation in traditional networks.

In this report, we investigate the value of data in understanding and quantifying losses; thus focusing on three of the four categories identified above. In WP3 we will use these findings and study the impacts of increased uptake of smart grid technologies in future networks, through a modelling framework that will enable improved estimation of losses taking into account network topology, data available, data quality and type. In the latest stages of the project we will use this to study losses DSO environment considering intentional operating decisions taken by the Operator as well as unintentional ones originating e.g. from market operations.

For the analysis of this WP we proceeded in two stages, first gathering and analysing network data from multiple sources, including SCADA, project data, representative profiles, and forecasted data, and secondly performing sensitivity analyses using a first iteration of our modelling framework, in order to capture:

- the value of data, including the amount, sampling rate, quality, and accuracy
- the effects of network topology considering both urban and rural configurations
- the effects of measurement location and the level of data aggregation

2 Methodology

This report contains a number of sensitivity analyses, investigating the impact of various factors on the accuracy with which network losses can be estimated. In each case, a set of observed network loading data – which will be treated as the ground truth for the purposes of the sensitivity analysis – was available. Some changes were made to that set of data, to represent a way in which the data could be observed in a given set of conditions, for example by reducing the sampling rate or adding an error onto the data. A load flow using a model of the local network was then performed using MATPOWER to assess the losses in the network using both the ground truth dataset, and the modified data set; the difference between the total losses in each case was then considered to be the estimation error introduced by a particular phenomenon. This process is illustrated in Figure 2.

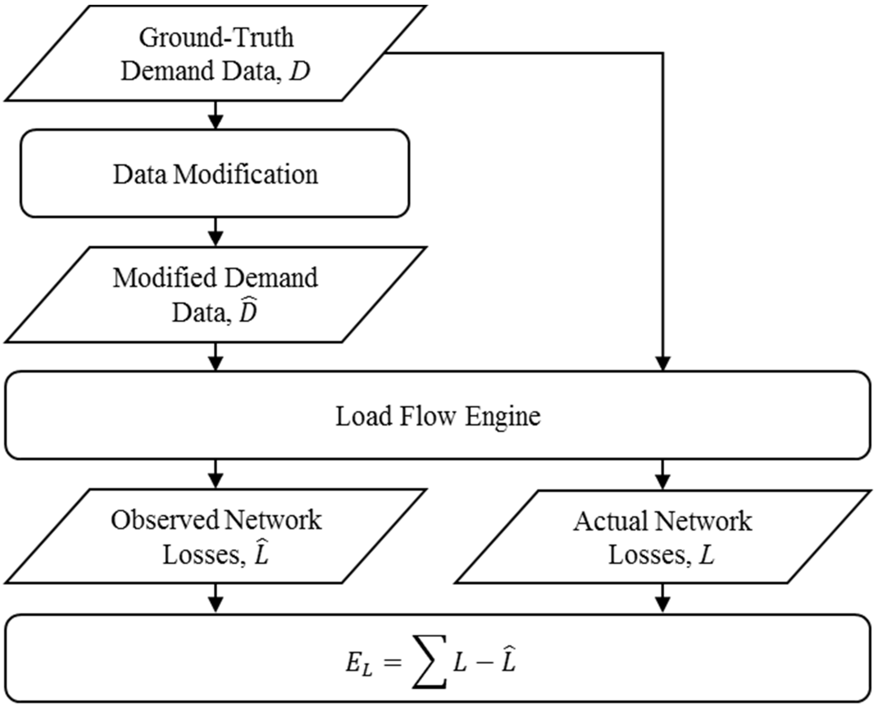

Figure 2: A flow chart showing the methodology used to carry out the sensitivity analyses.

The method is based on the assumption that the ground truth data set is free from errors, and can be used to calculate the network losses with perfect accuracy. While this will not be the case – the data used will have errors of all the types listed in this report – the results arising are still valid when each phenomenon is taken in isolation provided the ground truth data are *broadly representative* of the real behaviour of the system. The purpose of the method is to determine how a controlled change to some data affects the ability of an engineer to use those data to calculate network losses. Any errors in the data used as the ground truth can be disregarded, since this is being used as a baseline for adding the controlled errors, not to estimate the losses in the system those data actually represent.

3 Network and Consumer Datasets

The developed loss estimation method has the potential to use data inputs from multiple aggregation levels of the distribution network.

Whilst accuracy with regards to overall system loss determination is likely to increase as a function of increasing system data visibility, this could lead to significant increases in data requirements with minimal overall gain.

In order to inform future data requirements, a combination of consumer, and historic network monitoring data will be used to test the loss estimation model's sensitivity to receiving data from each of the potential network aggregation levels. These aggregation levels are most likely in line with typical voltage level segregations as follows:

3.1 LV Datasets

3.1.1 Consumer Data

Datasets available for testing include smart metering data from the Customer-Led Netwrok Revolution (CLNR) at a half hourly resolution, and Activating Customer Engagement (ACE) projects at a 20 second resolution, with the potential to resample at any given rate above this in addition to other datasets if necessary from relevant projects such as Irish Trials / EDRP datasets.

Data from the CLNR project covers a period of around 2 years, and includes consumer data for around 9000 normal tariff domestic customers, Small and Medium Enterprise (SME) consumers and Industrial and Commercial (I&C) consumer data from a number of sectors.

In addition to I&C customer demand data, there is also data from distributed generation customers connected to the distribution network at half hourly resolution, including wind farms, a limited number of hydro and landfill gas generators.

In addition to the standard tariff consumers, data regarding domestic consumers with various forms of embedded generation including air source heat pumps and solar photovoltaics is also available to potentially model such LCT adopting customers.

Generic consumer data is also available in the form of historical profiling data from Elexon. These are in the form of representative demand profiles for each of the consumer demand classes 1-8. These profiles are for each of the five yearly subdivision periods (Winter, Spring, Autumn, Summer and High Summer) at a half hourly resolution. These 8 demand classes cover a range of consumer types, from domestic and SMEs to a range of maximum demand I&C consumers.

There is the additional possibility to synthesise representative demand profiles with a high degree of disaggregation information through the use of a demand simulation method such as the Loughborough University's Centre for Renewable Energy Systems Technology's (CREST) model.

3.1.2 Network Data

Newcastle has access to a range of network data from a number of sources. Data from around 30 Northern Powergrid substations are available from iHost monitoring equipment. Historical data from a set of similar sites is also available from the CLNR project covering around 10 LV secondary substation sites, including LV distribution board data for outgoing downstream LV feeders, transformer monitoring data, ambient temperatures and at some sites additional data from link boxes and LV connected OLTCs. The data within these datasets is at 1 minute resolution. Historic information regarding customer composition within these networks is also available.

Similar datasets are also available for the Wellington Science secondary substation on the Newcastle Helix network, in addition to appropriate network models.

Northern Powergrid's Element Energy model gives output data for a range of future demand scenarios. This provides secondary substation demand characteristics for any secondary substation within Northern Powergrid's region for use within the losses modelling approach.

3.2 HV Datasets

The majority of the available HV network data comes from the legacy CLNR network monitoring installations. This consists of historic SCADA data for downstream HV feeders at the Rise Carr and Denwick Primary Substations (1 minute and Half Hourly resolution). This also includes transformer monitoring data including tap position and various temperature values including ambient.

There is also additional network monitoring data from HV installed RTTR equipment on overhead lines (OHLs) (Denwick only) and underground cables (UGCs) (Rise Carr only). This monitoring equipment provided a range of ambient meteorological parameters which can be used in conjunction with the electrical demands for temperature modelling and correlation if necessary. These monitoring sites also provided information on feeder currents and various equipment thermal parameters.

3.3 EHV Datasets

Datasets from the EHV system at present are limited to parameters measured at the RTTR OHL monitoring points on the incoming 66kV circuit to the Denwick Primary substation. This dataset provides the same information as found at the HV monitoring locations.

4 Sensitivity Analysis Studies

4.1 Time Resolution

This section examines the impact of time resolution on loss calculation at different voltage levels. Data of different granularity are employed in two feeders: 1) Alnwick Estates Teed feeder, part of the Denwick Primary Substation, and 2) Feeder 3 of the Maltby network. The first one is a high voltage (HV) rural network and the second one is a low voltage (LV) urban network. The type of the network – specifically, the voltage level – is shown to play a significant role on loss calculation, considering different sampling rates for loading.

4.1.1 Alnwick Estates Teed Feeder (HV)

Figure 3 shows the loading of Alnwick Estates HV Teed feeder for a year [1]; the sampling rate is 5 minutes, and zero values have been removed from the time series. The number of available data – after this process – is 93,312, which corresponds to 324 days.

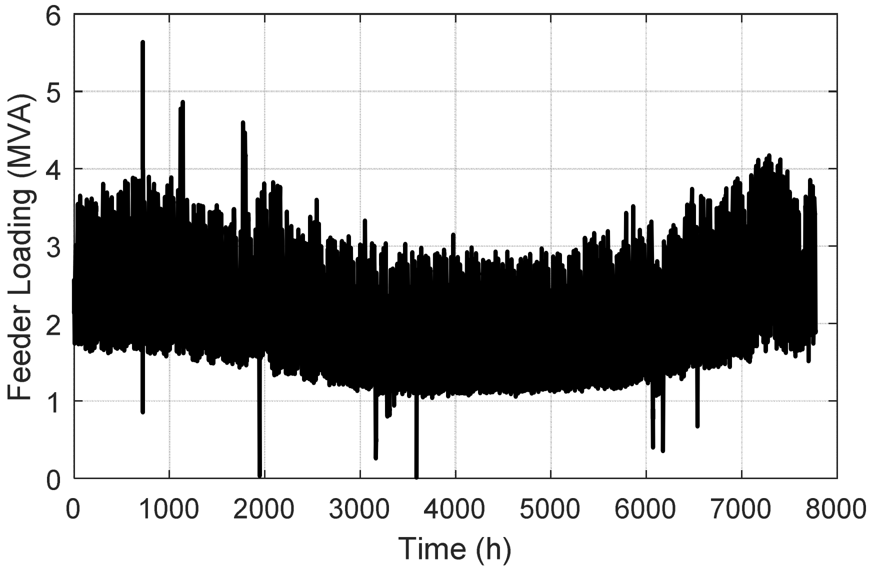

Figure 3: Alnwick Estates Teed feeder loading in 2014 (5-min resolution, after having removed zero values).

Figure 4 compares the loading of the feeder using 5-min, 30-min, and 1-h resolutions. It can be seen that the latter two granularities satisfactorily approximate the original profile. This is because of the higher load diversity at this voltage level, which leads to lower variability of the loading compared to low voltage, for example, as will be shown later.

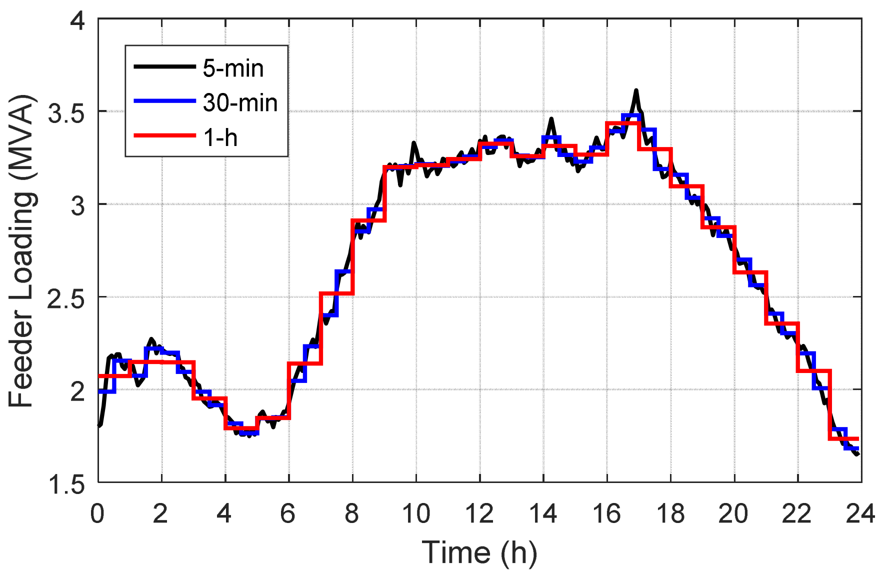

Figure 4: Comparison of feeder loading for a single day (5-min, 30-min, and 1-h resolutions) for Alnwick Estates Teed feeder.

Figure 5 compares the variation of losses using the three aforementioned sampling rates. Power loss profiles accentuate the variation of load, as can be seen in these figures, and the 30-min and 1-h resolutions provide a good approximation of the losses. The results of the simulation for the whole year are shown in Table 1. It should be noted that the loading profile of each load point of the feeder was considered to be equal to the feeder loading multiplied by the peak demand of the load point divided by the sum of the peak demand of all load points; this disaggregation method is also employed in sections 4.2 and 4.3.

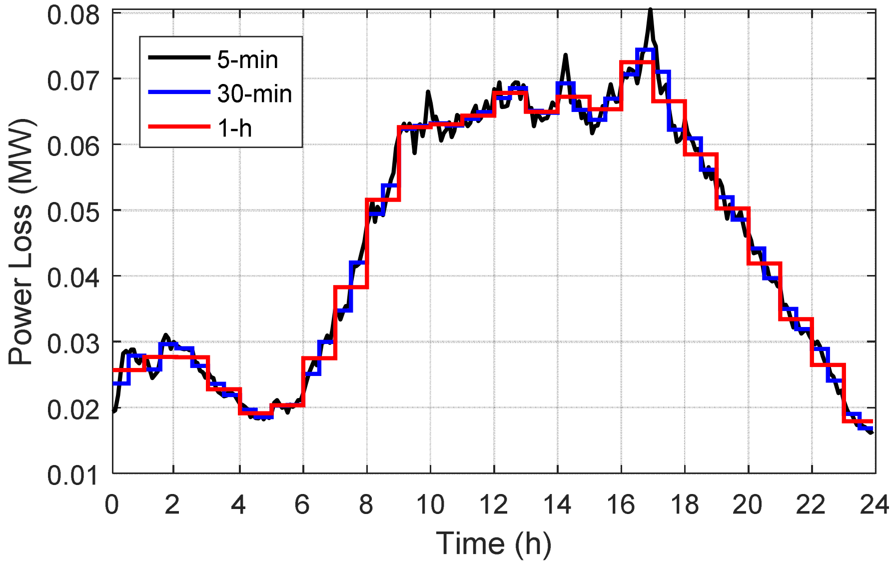

Figure 5: Power loss variation for a single day (5-min, 30-min, and 1-h resolutions) for Alnwick Estates Teed feeder.

Table 1: Results of one-year simulations using 5-min, 30-min, and 1-h time steps for loading at Alnwick Estates Teed feeder.

Time Resolution	Energy Losses (MWh)	Error (%)
5-min	276.86	–
30-min	276.69	0.06
1-h	276.47	0.14

Table 1 shows that the errors in loss estimation are very small. Although there are significant power loss errors at some individual time steps, most of them are small, and most importantly the sign of the error alternates, so the mean error is close to zero. However, this is enough to produce a small error, which increases linearly with the sampling rate. Figure 6 shows the loss estimation error graphically, with a linear trend line fitted. The loss is small in all cases, and extrapolating the trend line to zero suggests that the error arising from the 5-minute averaging (the highest sampling rate available) is only around 0.01%. These low errors are a result of the low level of variability in the data (a result of the high level of diversity within the larger load groups found on HV networks). This means the averaged lower resolution data provide a good approximation to the 5-minute data.

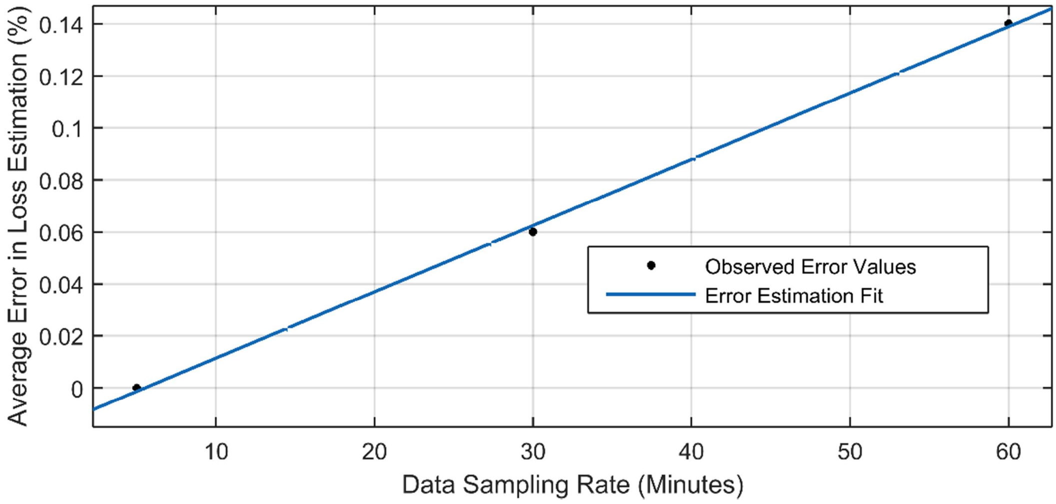

Figure 6: The relationship between sampling rate and loss estimation error for Alnwick Estates Teed feeder.

4.1.2 Maltby Network Feeder 3 (LV)

Figure 7 shows the loading of feeder 3 of Maltby LV network for a day in July 2012 using 1-min, 5-min, 30-min, and 1-h data sampling rates [1]. It is clear that half-hourly and hourly sampling rates do not provide as good approximations to the loading as in the previous section. Even when the 30-min and 1-hour data are compared to the 5-min data, the representation is still significantly worse than the same approximation at HV. This is because loading is more volatile at LV, where fewer customers lead to lower diversity and therefore a more variable profile. These differences in the time series are accentuated in power loss profiles – which are illustrated in Figure 8– especially at peak hours in the evening.

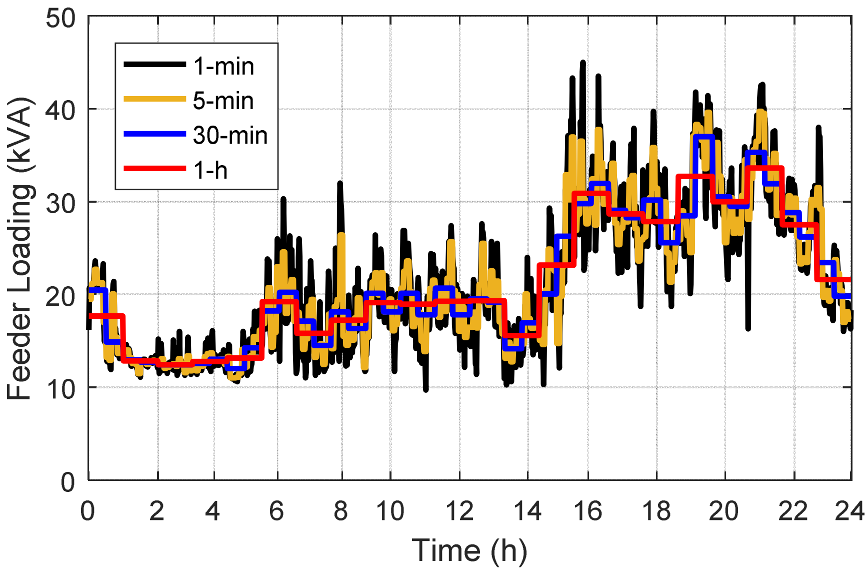

Figure 7: Comparison of feeder loading for a day, using 1-min, 5-min, 30-min, and 1-h resolutions at Maltby Network Feeder 3.

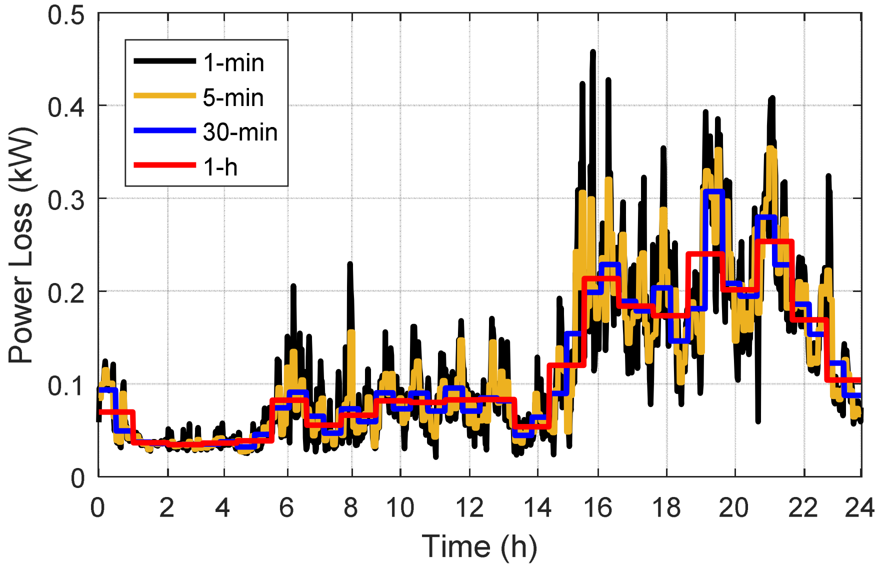

Figure 8: Variation of power losses during a day in July, 2012, using 1-min and 1-h sampling rates at Maltby Network Feeder 3.

Table 2 shows that the errors have more significant values compared to those in the HV case, due to the increase in loading variability.

Table 2: Results of one-day simulations using 1-min, 30-min, and 1-h time steps for loading at Maltby Network Feeder 3.

Time Resolution	Energy Losses (kWh)	Error (%)
1-min	2.54	–
5-min	2.51	1.06
30-min	2.47	2.63
1-h	2.46	3.18

Figure 9 shows the relationship between data sampling rate and loss estimation error for the LV test case. The error using a 5-min resolution has been set to zero to allow comparison with the HV case. Unlike the HV case, the trend is not linear, and the fitting suggests that the additional error arising from the 1-minute sampling is around 1.5%.

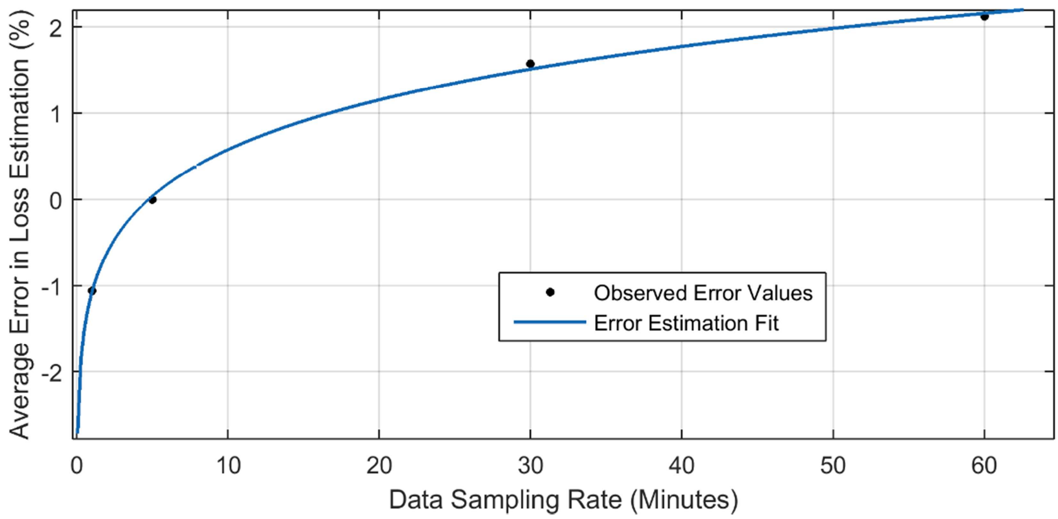

Figure 9: The change in loss estimation error with sampling rate for Maltby Network Feeder 3.

4.1.3 Conclusions

This section studied the effect of loading sampling rate on loss evaluation using 1-min, 5-min, 30-min, and 1-h time steps. Two networks – an HV rural and an LV urban – were considered in the analysis, and the results indicated that the most significant factor that impacts loss estimation (in terms of time resolution) is the variability of the feeder loading, which is related to the load diversity; in general, a larger load group will have a higher level of diversity, and therefore less variability, which means at HV load variability is generally lower than at LV. The error in energy losses for a 1 hour resolution was approximately 0.1% and 3% in HV and LV, respectively. Increasing the sampling rate leads to greater underestimation of energy losses because of the resulting smoother profile. These results are in agreement with the findings from a similar study undertaken by the University of Sheffield [2]; in that study, the authors focus on LV customers, and used aggregated smart meter data to study demand groups of different sizes. The loss estimation error observed for the LV feeders in this report (with a peak demand of 40kVA) is very comparable to the errors found by the Sheffield study when aggregating 25 smart meter customers.

4.2 Missing Values

4.2.1 Data unavailability modelling and simulations

This section analyzes the effect of missing values on the calculation of energy losses. Measuring devices are imperfect, and it is common to find missing or bad data; this phenomena is modelled through a two-state continuous Markov process [3], which is illustrated in Figure 10. The frequency and duration of the sensor failures are represented by the Mean Time to Failure (MTTF) and Mean Time to Repair (MTTR) respectively. The MTTF represents the average time for which the sensor will operate without a failure, and the MTTR represents the average duration for which it will not operate. In both cases, different combinations MTTF and MTTR give various data unavailabilities. Figure 11 shows an illustrative example of an operating sequence of a measuring device, using an MTTF of 1 week and an MTTR of 1 day. Missing values are filled using linear interpolation – a standard, widely used method – between the nearest available data, and energy losses are compared to the value corresponding to the complete dataset.

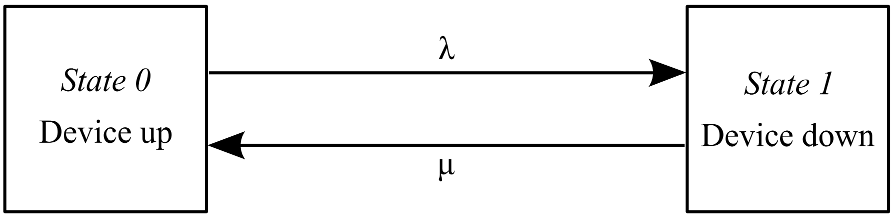

Figure 10: Measuring device state space diagram (λ and μ represent the failure and repair rates, respectively).

Sections 4.2 and 4.3 use simulations of Haxby Road primary distribution network (see Figure 12), which is an HV urban distribution network comprising 7 primary feeders and 56 load points. Figure 13 shows the power loss time series for all 7 feeders of the network over a year), using 17,520 half-hourly data, with zero missing values from the dataset. The overall energy losses that correspond to Figure 13 – which are considered the base case – are 493.89 MWh.

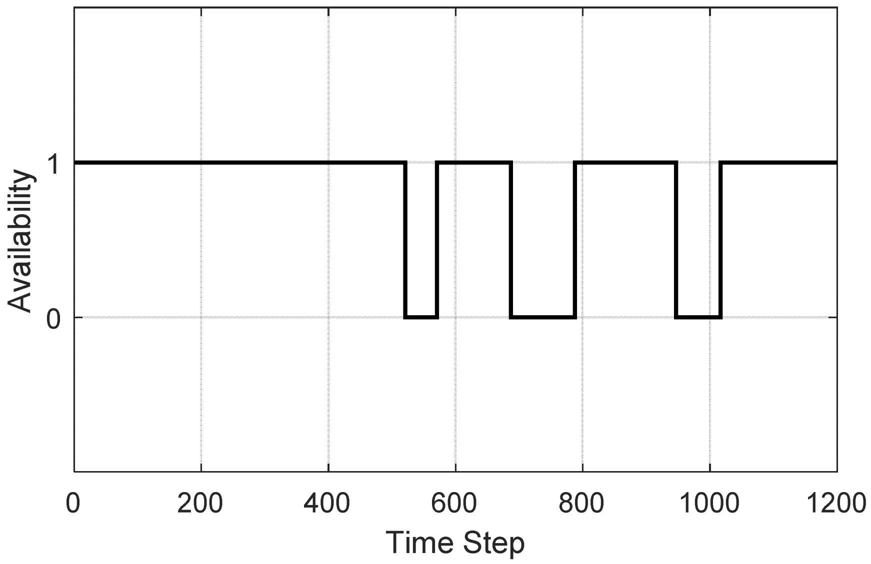

Figure 11: Simulated operating sequence for a measuring device using an MTTF of 1 week and an MTTR of 1 day; illustrative example.. Each time step is 30 minutes.

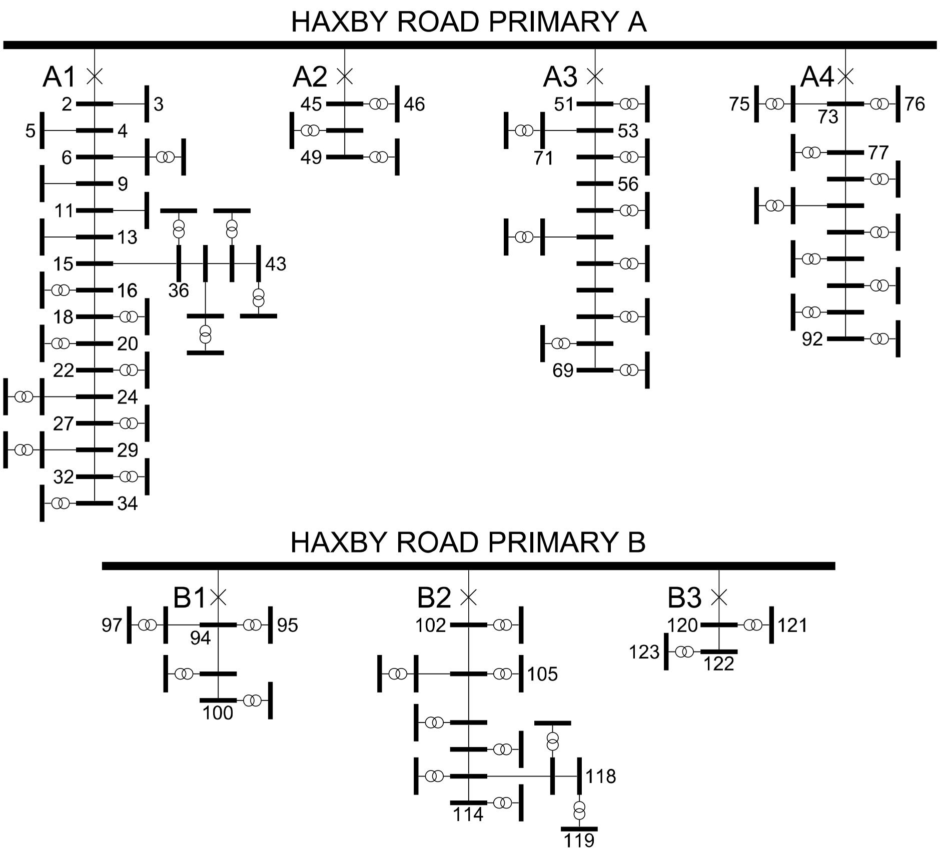

Figure 12: Haxby Road T2 T3 Primary distribution network diagram.

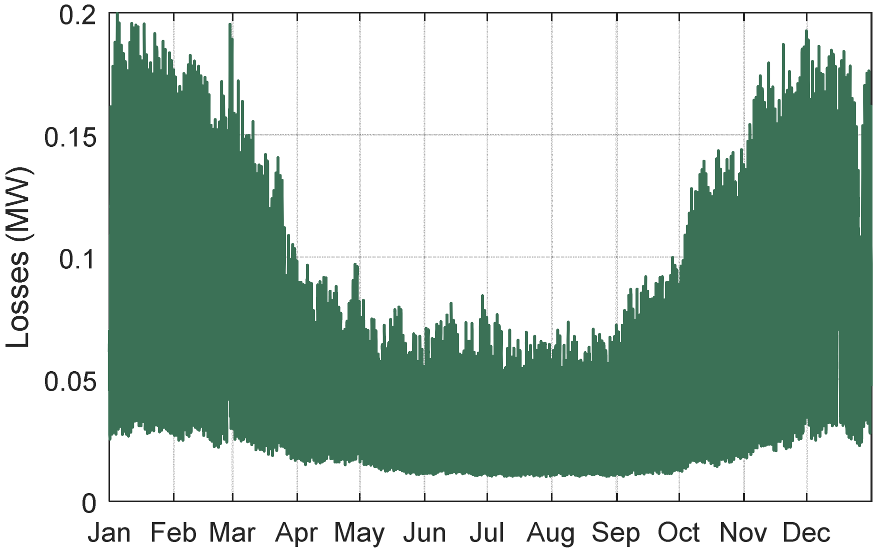

Figure 13: Modelled power loss time series for Haxby network over a year (2017), considering no missing data; overall energy losses = 493.89 MWh. Each time step is 30 minutes.

Figure 14 illustrates the simulation approach over a 25 day period. The original profile (with no missing data) is shown in blue; the availability of the measuring device is shown in black; and when the measurement is unavailable, the loading is approximated via linear interpolation based on the two nearest available values, which is shown using the red line. The process described is carried out for one year, which constitutes one run of a Monte Carlo simulation; 1000 runs are performed for each combination of MTTF and MTTR which allows the long-term effect of these reliability parameters to be investigated; numerous combinations of these are considered to explore the impact of missing data on energy loss estimation accuracy. The results of these simulations are shown in Tables 3 and 4. Unavailability (U) – in terms of MTTF and MTTR – is given in (1).

$$U = \frac{MTTR}{MTTF + MTTR} \tag{1}$$

Time to failure (TTF) and time to repair (TTR) for the Monte Carlo simulations are considered to follow exponential distribution with mean values $MTTF$ and $MTTR$, respectively.

$$TTF(x) = -MTTF \cdot \ln(x) \tag{2}$$

$$TTR(y) = -MTTR \cdot \ln(y) \tag{3}$$

where x and y are two random numbers (0, 1).

It is clear from Tables 3 and 4 that the greater the amount of unavailable data, the greater the underestimation of losses (represented by mean) and the associated uncertainty (expressed through standard deviation). The underestimation of losses occurs because linear interpolation decreases the variability of the original time series. Moreover, MTTR is shown to have a more substantial effect on the uncertainty of loss estimation than MTTF, i.e. for the same unavailability, the case with the higher MTTR, leads to a higher standard deviation.

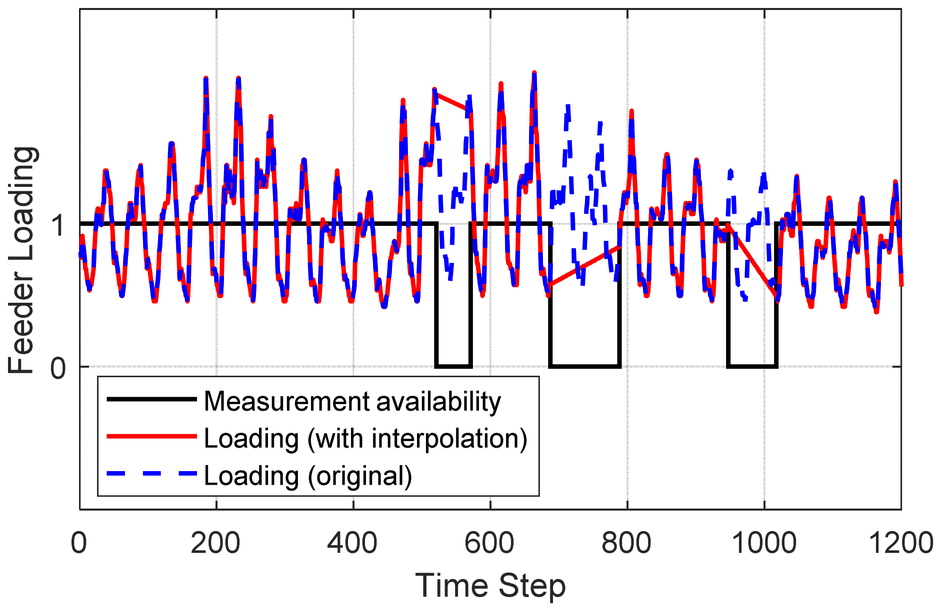

Figure 14: Illustrative part of a simulation (25 days) carried out to investigate the effect of data unavailability on energy losses. Each time step is 30 minutes.

Table 3: Results of 1000 Monte Carlo simulations with an MTTF of 1 week and varying MTTR. Mean and standard deviation refer to total energy losses in MWh for a year; losses with no missing data are 493.89 MWh.

MTTR	6 hours	12 hours	1 day	2 days	3 days
U (proportion of unavailable data)	3.45%	6.67%	12.50%	22.22%	30.00%
Mean (MWh)	493.44	493.05	492.50	491.55	490.35
Error (MWh)	0.45	0.84	1.39	2.34	3.54
Error (%)	0.09	0.17	0.28	0.47	0.72
St. dev.	0.70	1.36	2.54	5.20	6.89

Table 4: Results of 1000 Monte Carlo simulations with an MTTR of 1 day and varying MTTF.

MTTF	1 month	2 weeks	10 days	1 week	3 days	1 day
U (Proportion of unavailable data)	3.23%	6.67%	9.09%	12.50%	25.00%	50.00%
Mean (MWh)	493.51	493.12	492.82	492.50	490.87	488.28
Error (MWh)	0.38	0.77	1.07	1.39	3.02	5.61
Error (%)	0.08	0.16	0.22	0.28	0.61	1.14
St. dev.	1.30	1.93	2.30	2.54	3.62	6.15

Figure 15 shows the relationship between MTTR and the average loss estimation error for a fixed MTTF of 1 week. The estimation error increases linearly with the MTTR of the sensor. The adjusted R^2 value of the fit – an indication of how well the fit represents the data – is 0.9953 against a maximum value of 1.

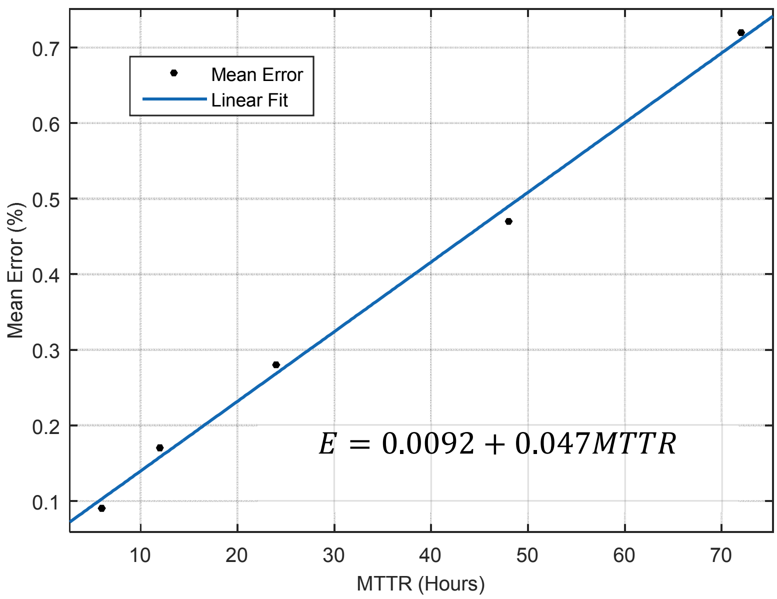

Figure 15: The relationship between MTTR and loss estimation error with a fixed MTTF of 1 week

Figure 16 shows how the estimation error varies with respect to the MTTF; in this case the MTTR was fixed at 1 day. In this figure, the trend follows a rational relationship, with very low MTTF values having a comparatively high impact on estimation error, but sensor failures with a MTTF of more than 14 days providing little improvement. The adjusted R^2 of this fitting is 0.9988, again implying the fit is an extremely good representation of the relationship.

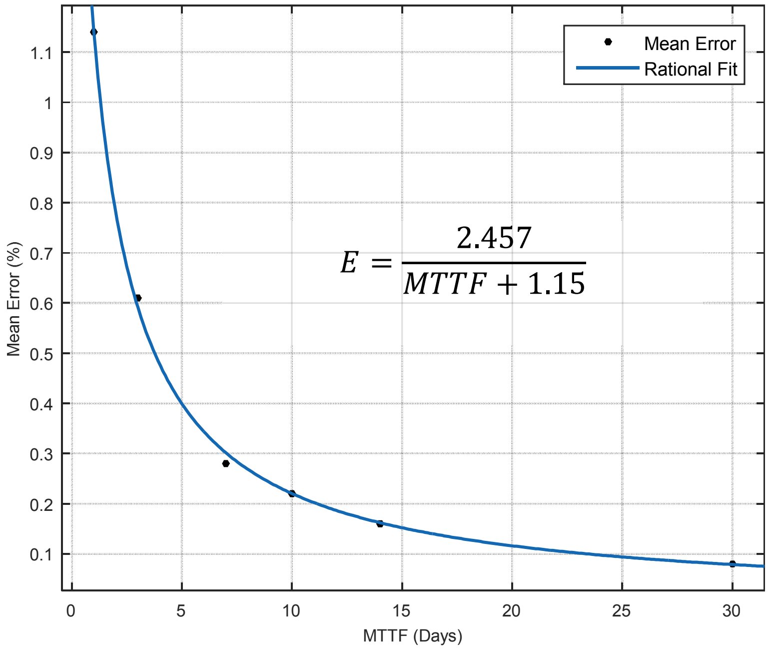

Figure 16: The relationship between estimation error and MTTF with a fixed MTTR value of 1 day

4.2.2 Conclusions

The impact of data unavailability on estimation of energy losses was examined in this section and was found that the error is relatively low compared to the amount of missing data – 45 (equivalent full) days with no available data led to a maximum error of 1.5%. More missing data resulted in greater underestimation of energy losses and higher uncertainty as well. Finally, extended periods of missing values compared to more frequent (but shorter) unavailability periods (with the same ratio of available data) play a more important role in loss estimation in terms of uncertainty. The error estimation increased linearly with the duration of the periods of missing data (represented by a MTTR value), and had a rational ($\frac{1}{x}$ type) relationship with the spacing between periods of missing data, represented by the MTTF. The adjusted R^2 value for both of these relationships was greater than 0.99.

4.3 Uncertainty of load and network parameters

This section analyses the effect of the uncertainty regarding load and network parameters; random and systematic errors are considered, and correlation is examined as well.

4.3.1 Load uncertainty

Load uncertainty is analysed by employing Monte Carlo simulation; equation (4) shows how the real power of load point k for simulation i at time t is derived. Random errors are considered by modifying the standard deviation (σ) of normally distributed random numbers (r), and systematic errors by adjusting their mean (μ).

$$P_{i,k}(t) = D(t) \cdot a_k \cdot (1 + r_i(\mu, \sigma)) \quad (4)$$

where $P_{i,k}(t)$ is the original observed power at load point k for simulation i at time t; $D(t)$ is the demand of the corresponding feeder; a_k is the ratio of the peak demand of the load point to the sum of peak demands of all load points of the feeder; and $r_i(\mu,\sigma)$ is the value of a normally distributed random number with mean μ and standard deviation σ, at simulation i.

Figure 17 and Figure 18 present the demand quantiles (i.e. the probability of demand at each time step) of 100 simulations of (4) for load point 3 of Bowling Green Court feeder of Haxby Road network, which is used for the analysis in this section. Figure 17 corresponds to μ = 0 and σ = 0.1, while μ = 0.1 and σ = 0.1 have been used for Figure 18. In the former figure the simulated profiles vary around the original loading profile, whereas in the latter figure, the simulated time series are noticeably shifted up; this is because μ has a non-zero value, which represents a systematic error.

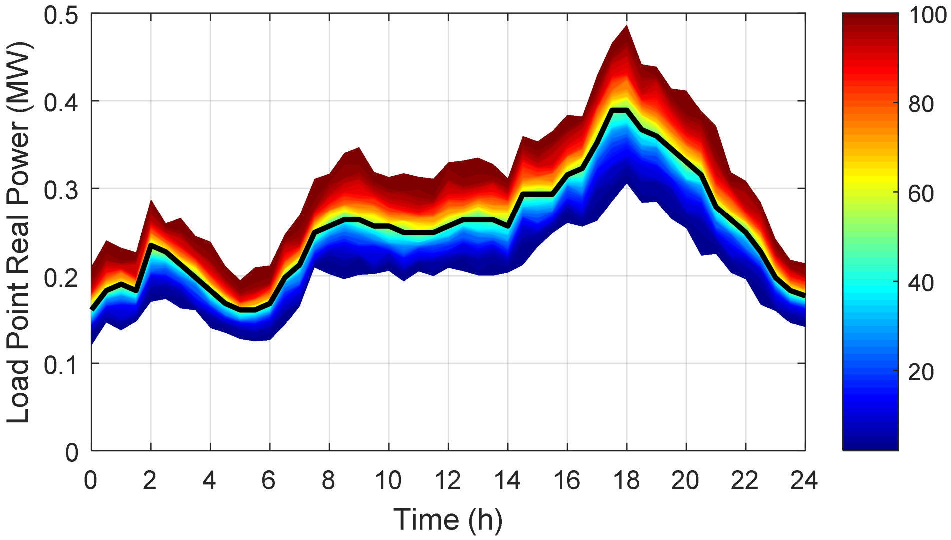

Figure 17: Demand quantiles of the load profile of load point 3 of Bowling Green Court feeder (100 simulations) using μ = 0 and σ = 0.1 for the normally distributed random number; original load profile in black.

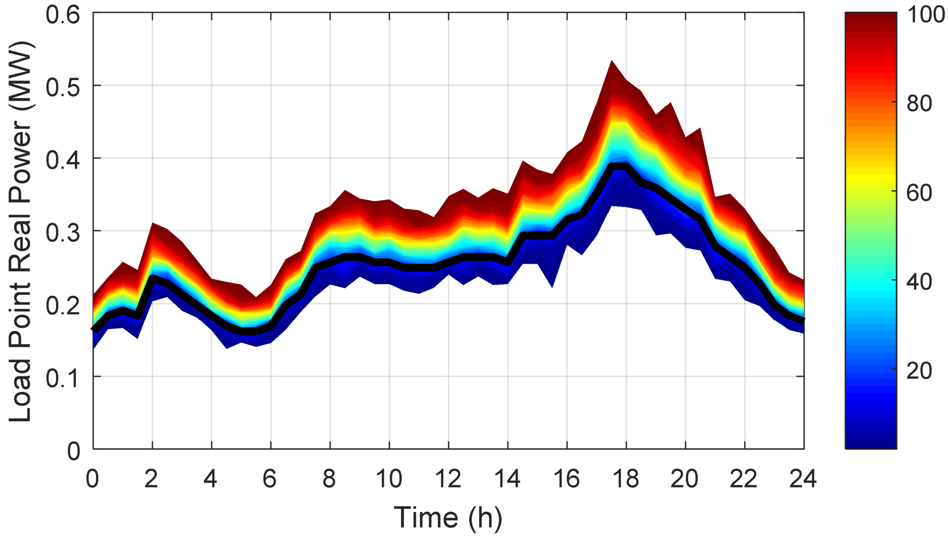

Figure 18: Demand quantiles of the load profile of load point 3 of Bowling Green Court feeder (100 simulations) using μ = 0.1 and σ = 0.1 for the normally distributed random number; original load profile in black.

The mean value of energy losses for the cases above (for the whole network) – having run 100 Monte Carlo simulations for each one – are 496.48 MWh (μ = 0, σ = 0.1) and 602.62 MWh (μ = 0.1, σ = 0.1), as shown in Figure 19. The energy losses corresponding to μ = 0.1 and σ = 0, i.e. considering only systematic error, are 600.1 MWh; this means that the existence of systematic error is more significant than that of a random one. The existence of a random error adds some variability around the original time series (see Figure 17) and can be used to reflect the stochasticity of load during operation (e.g. [4]); this variability increases (the estimation of) losses compared to the case that neglects uncertainty. However, this increase is lower than that of a systematic error, for which the whole time series is shifted, and therefore the losses are increased to a greater extent. Figure 19 presents a sensitivity analysis of load uncertainty (in terms of random and systematic error) on estimated energy losses for a year for the whole network; this can be used to assess losses at different levels (and combinations) of demand uncertainty.

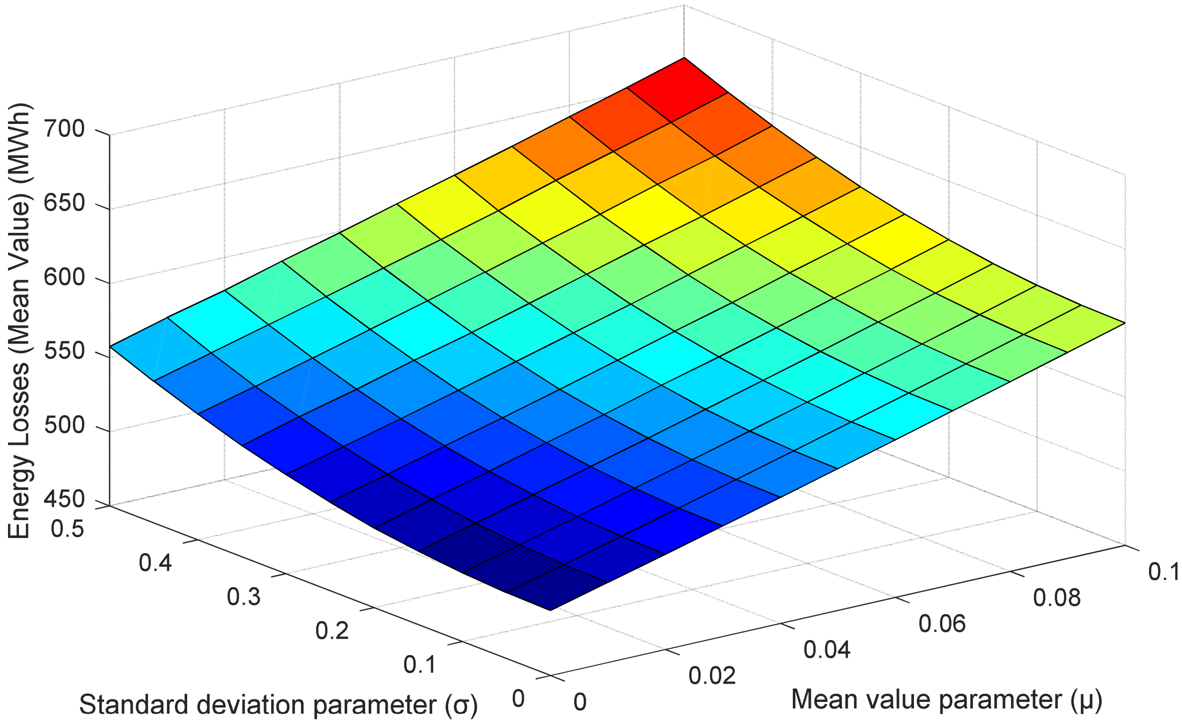

Figure 19: Surface plot of mean energy losses for different combinations of random and systematic errors; base case (μ = 0, σ = 0) energy losses = 493.89 MWh.

4.3.2 Measurement Accuracy on Real Networks

The potential measurement errors encountered on real networks are a compound value encompassing the error from several discrete processes, namely:

1. Current Transformers (CTs) and Voltage Transformers (VTs) are used to transform the voltage and current on the real network to lower values which can be safely measured by a transducer. These devices have a ratio which is guaranteed to a certain accuracy depending on the *accuracy class* of the CT or VT. In the majority of installations, the CTs and VTs have accuracy class 1 [5], which means the ratio is accurate to ±1%; a systematic error of up to 1% can be introduced. The accuracy of the CT is also affected by its power factor and burden, which can lead to a random error of up to ±1%.
2. A transducer on the secondary coil of the CTs and VTs is used to measure the current or voltage and pass this data onto a relay. The relays used by Northern Powergrid's RTUs which provide data to Pi have a current measurement error of 1-2%.

3. The measurement from the CTs and VTs are then scaled up to represent the real quantity.
4. This scaled analogue measurement is then converted to a 7-bit digital signal – this introduces a quantization error. Assuming the 7 bit signal is used to represent a value from 0-150% of the transformer rating, this will introduce an error of $\frac{150}{7^2} = 1.17\%$

Without full knowledge of the exact natures of all the errors taking place, a reasonable approximation to the sensor errors in the real system would be a systematic error of 1% - arising from the CT – and a random error of 4.2% - combining the transducer, scaling, and quantization errors. This corresponds to an underestimation of losses by around 10 MWh/year for the study network, or around 2%.

4.3.3 Network parameter uncertainty

This subsection deals with the uncertainty of network parameters (resistances and reactances). The uncertainty is introduced in the same way as in the previous section (see (4)) and 100 Monte Carlo simulations are carried out; then the mean and the standard deviation of total annual energy losses are calculated.

$$\begin{aligned} R_{i,j} &= R_j \cdot (1 + r_i(\mu,\sigma)), \\ X_{i,j} &= X_j \cdot (1 + r_i'(\mu,\sigma)) \end{aligned} \tag{5}$$

where $R_{i,j}$, $X_{i,j}$ are the resistance and reactance of branch j for simulation i, respectively; R_j, X_j are the original resistance and reactance of branch j, respectively; and $r_i(\mu,\sigma)$, $r'_i(\mu,\sigma)$ are the values of two normally distributed random numbers with mean μ and standard deviation σ, at simulation i.

Figure 20 and Figure 21 show the mean value and standard deviation of energy losses, respectively, for different combinations of network parameter uncertainty, in terms of random and systematic error. It can be seen that the mean value of energy losses depends only on the value of systematic error (see Figure 20), and the standard deviation of energy losses is influenced only by the random error (see Figure 21).

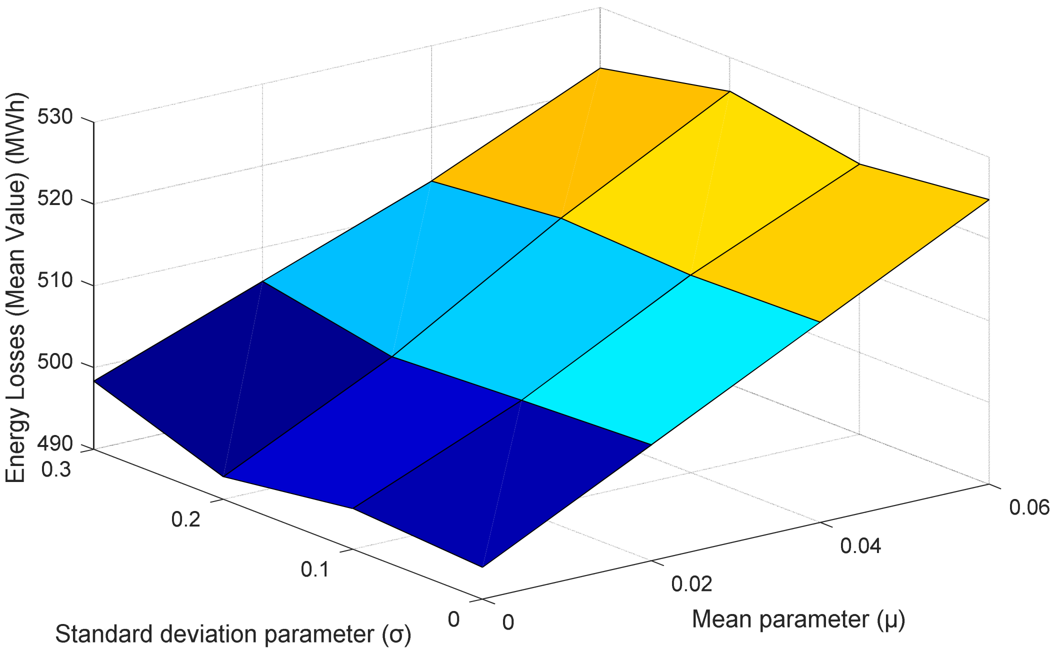

Figure 20: Sensitivity analysis of network parameter uncertainty (random and systematic error) on the mean value of energy losses.

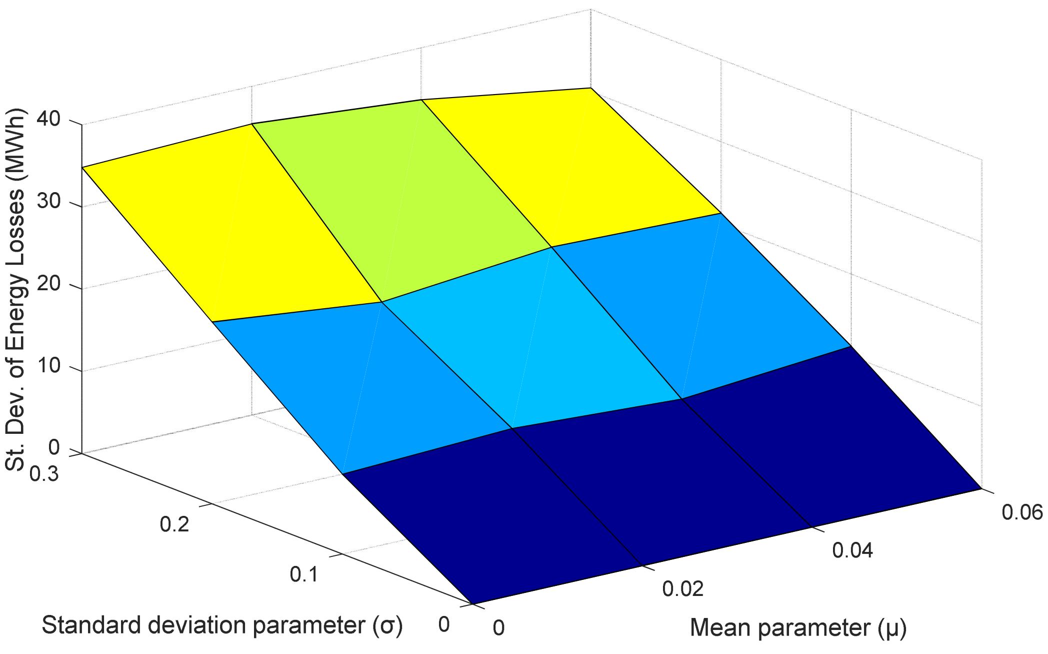

Figure 21: Sensitivity analysis of network parameter uncertainty (random and systematic error) on the standard deviation of energy losses.

Figure 22 illustrates the probability distribution of energy losses for a specific combination of systematic and random errors, namely $\mu = 0$ and $\sigma = 0.3$. The mean value of energy losses is very close to the base case value (i.e. 493.89 MWh), because of the zero systematic error; the standard deviation of energy losses is 34.81 MWh (which corresponds to 7% of the mean value), which is low compared to the magnitude of the input random error.

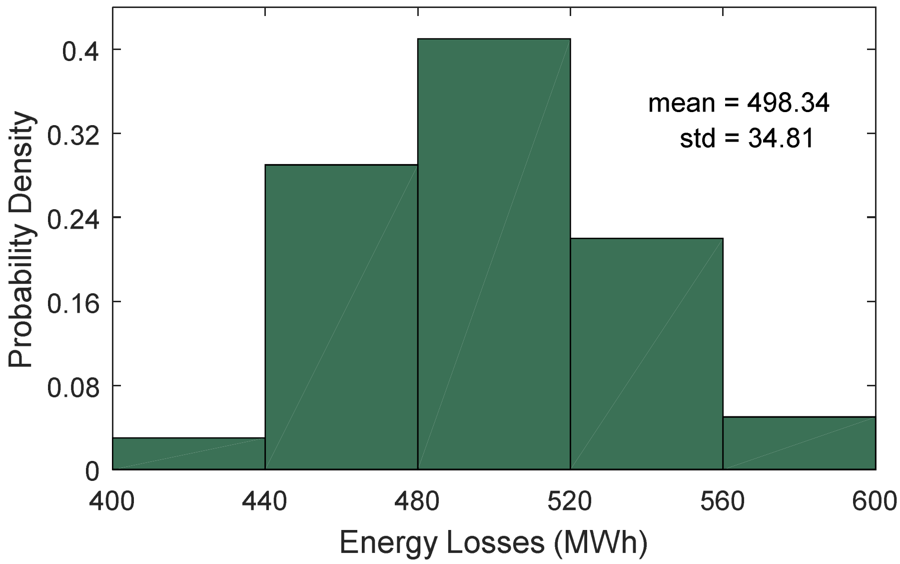

Figure 22: Probability distribution of energy losses for $\mu = 0$ and $\sigma = 0.3$.

These results suggest that a systematic error in the network parameters will lead to a significant over or under estimation, while a random error in network parameter values will affect the range of values that the true value of the losses could take. Given that the parameters studied are affected by drivers which could affect all conductors in a given area (for example, air temperature and wind speed affecting overhead conductor temperature), a systematic set of network parameter errors is not unlikely to occur.

4.3.4 Load correlation

The loads in a given distribution network will be affected by many of the same factors – for example low temperatures, or sunset – but the extent to which this is true will vary from case to case. For example, a feeder with a significant amount of industrial and commercial demand will behave very differently to one with only domestic users. The extent to which the demands at two given load points are driven by the same external factors can be modelled by setting the correlation between different loads, and its impact on network losses is investigated in this section.

The correlation – with a correlation coefficient specified for each simulation – between the demands were created by using correlated strings of pseudo random numbers as inputs into the model for the uncertain load described in section 4.3.1. An example of how the magnitude and correlation of these values affected the demand profiles are shown in Figure 24 and Figure 25, with the original data provided in Figure 23, for reference. Note that a correlation coefficient closer to unity indicates a stronger correlation.

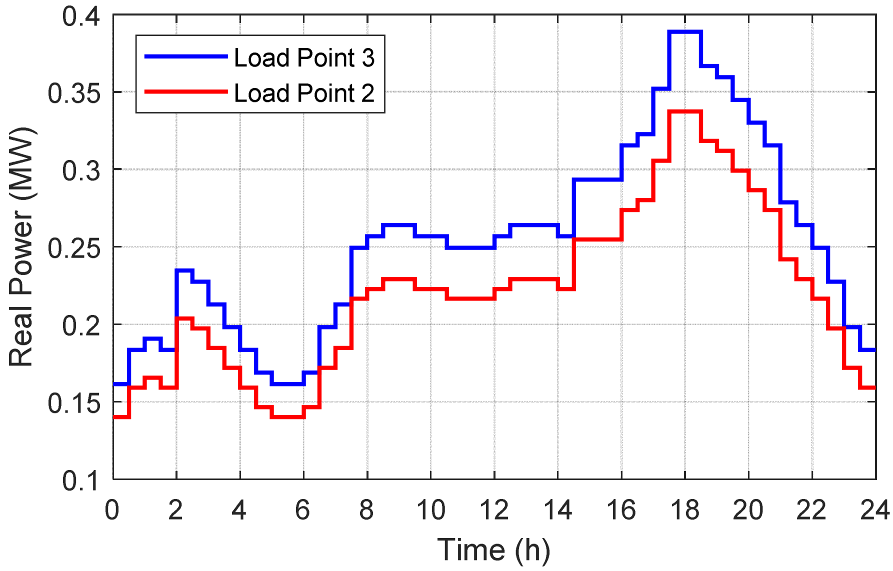

Figure 23: Original load profiles of load points 2 and 3 of Bowling Green Court feeder for a single day.

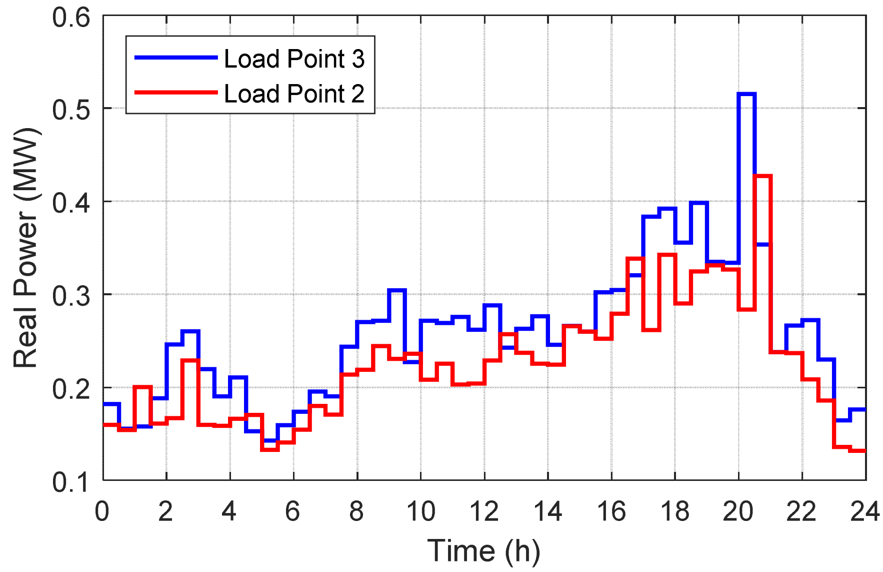

Figure 24: Synthesized load profiles of load points 2 and 3 of Bowling Green Court feeder, using $\mu = 0$, $\sigma = 0.1$, and zero correlation.

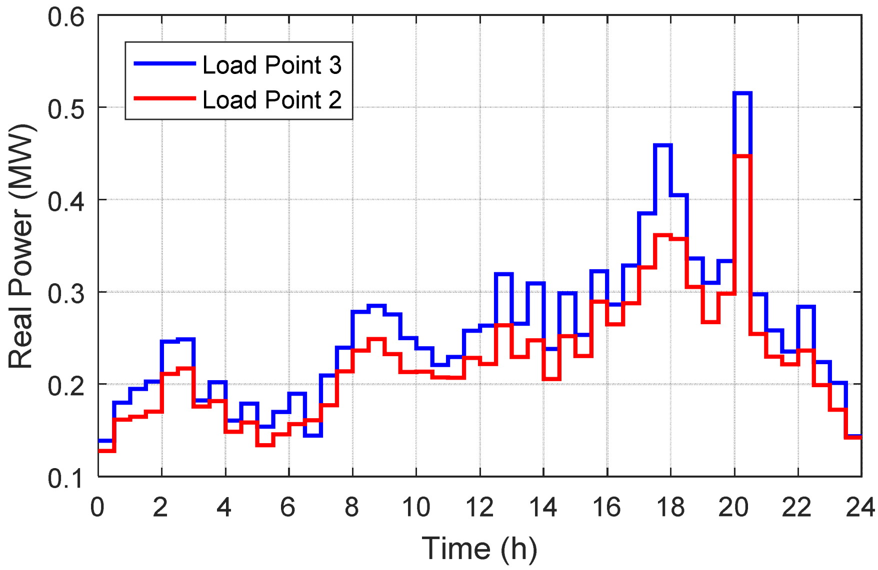

Figure 25: Synthesized load profiles of load points 2 and 3 of Bowling Green Court feeder, using μ = 0, σ = 0.1, and a correlation coefficient of 0.95.

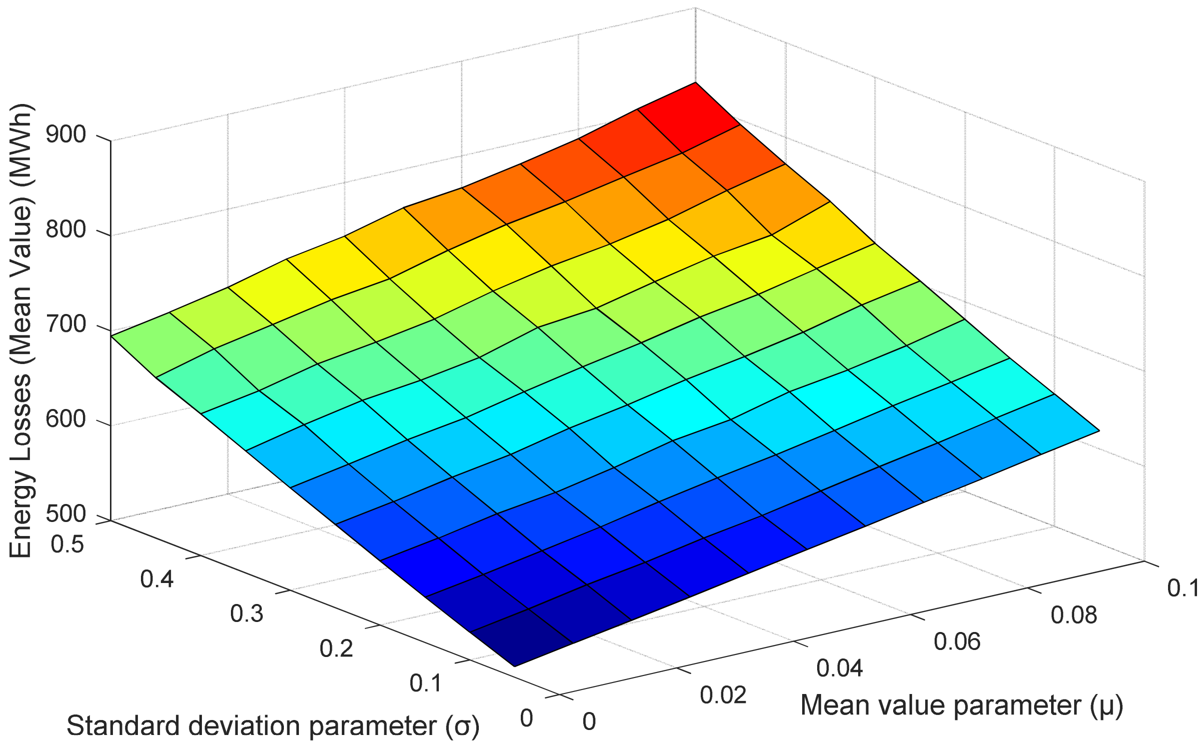

Figure 26: Surface plot of mean energy losses for different combinations of random and systematic errors (correlation coefficient = 0.6).

The resulting impact on loss estimation is shown in Figure 26 and Figure 27, which correspond to correlations between all the load points on the feeder of 0.6 and 0.8

respectively. These can also be compared to Figure 19, which shows the results of the same analysis with zero correlation between the demands.

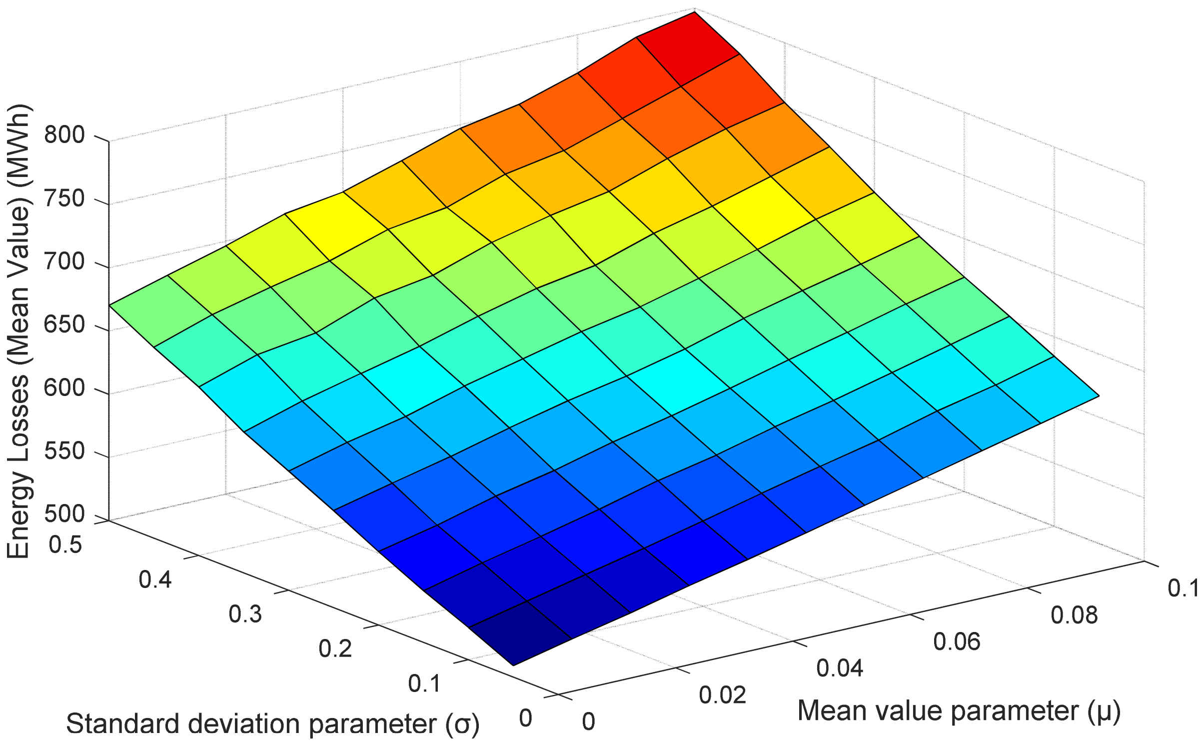

Figure 27: Surface plot of mean energy losses for different combinations of random and systematic errors (correlation coefficient = 0.8).

The results show that the correlated results give higher losses than if we assume the demands to be independent – this is because it is likely that more peaks will occur concurrently, leading to higher overall losses as a result of the Joule heating effect. What is less clear is the specific relationship between the correlation and the losses, since in some conditions, the highest correlation (0.8) actually yielded lower losses than the simulations with a correlation of 0.6. A possible explanation for this is that, as the correlation increases, the peaks start to occur simultaneously, which leads to higher losses; however, as the correlation becomes very high the level of variability – which has been clearly linked with higher losses – in the networks begins to fall, which causes losses to fall as well.

4.3.5 Conclusions

Uncertainty in demand observations leads to an underestimation of network losses in most cases; if there is a systematic error in the demand measurement, this has a more significant impact, though this could be more easily corrected through recalibration of the equipment.

The error arising from a random error was less significant, but is also harder to eliminate. Correlation between demand groups will lead to an increase in the losses overall, but the exact nature of this relationship will require further investigation.

5 Impact on Future Modelling

Future work within this project will focus on estimation and prediction of network losses in present and future network scenarios. These estimations will be based on data, with similar characteristics as those explored in the sensitivity analyses in this report. The results clearly show that if these effects are not included within the analysis then it is likely that the losses will be, on average underestimated and will have a higher level of uncertainty – a given observation could indicate a wider range of true values.

With further analysis, a set of adjustment factors or tables could be created, which would allow an engineer to easily include the impact of the data errors into their calculations when estimating network losses. These tables would require analysis of a greater variety of data sets and networks, to investigate the variation in the sensitivities calculated in this report.

5.1 Losses in the DSO environment

The results presented in this report have some implications for a future DSO environment. The key outcome from all of the results presented is that higher variability within the network demand leads to higher losses. In the DSO environment, a greater number of assets acting in new and unconventional ways could add to this variability, and therefore lead to an increase in losses for the same level of network utilisation. However, any DSO service which aims to reduce the peak demand through load shifting will lead to a reduction in variability, and therefore a reduction in losses. It was also shown that loss estimation is more accurate at higher voltage levels for the same level of parameter uncertainty, therefore DSO actions which are targeted at reducing losses in the HV and EHV networks have a higher chance of meeting their goals than those directed at the LV network.

6 Conclusion

This report has investigated a number of factors which affect the estimation of network losses, specifically data resolution, missing data, data uncertainty, and the correlation between demand values. In many cases, relationships between these parameters and the resulting error in loss estimation could be calculated via sensitivity analysis, which can lead to improved loss estimation methods. In other cases, the specific nature of the relationship may require more detailed studies.

More specifically it was clearly shown that higher variability within the network demand leads to higher errors in loss estimation and specifically underestimation. As load variability is much lower in HV, because of the large number of customers supplied relative to LV, loss estimation errors will tend to be lower. Reducing the data sampling rate leads to greater underestimation of energy losses because of the resulting smoother profile, yielding the conclusion that data sampling has a more profound role to play in loss estimation in LV networks.

Regarding the impact of data unavailability on estimation of energy losses, it was found that the error is relatively low compared to the amount of missing data. It was also shown that extended periods of data unavailability had a more profound effect in errors in loss estimation than shorter but more frequent periods. The error estimation increased linearly with the duration of the periods of missing data, and had a rational ($\frac{1}{x}$ type) relationship with the spacing between periods of missing data. This suggests that long periods of measuring device failures will have a more significant impact in loss estimation than unavailability due to shorter interruptions such as communication failures. In every case more missing data resulted in greater underestimation of energy losses.

Uncertainty in load measurement was also shown to result in underestimation of network losses for the majority of the cases explored. Systematic errors, which could for example arise from calibration of measuring equipment or conversion of analogue into a digital values, have a more significant impact in loss estimation. Random errors were shown to be less significant, but could be harder to eliminate. It was also found that correlation between demand groups will lead to an increase in the loss estimation overall, but the exact nature of this relationship will require further investigation.

Besides the learnings derived from these observations, for a significant number of the studies undertaken, generalizable mathematical relationships were established between the various parameters considered as variables in the analyses, and loss estimation. These could be integrated with loss estimation models and other method for calculating losses in order to produce more accurate results. The studies have also provided clear evidence around the importance of errors originating from measuring equipment and quantified their effect on network loss estimation for a variety of cases.

7 References

[1] *Customer-Led Network Revolution*. Available: http://www.networkrevolution.co.uk/

[2] G. Poursharif, A. Brint, M. Black, and M. Marshall, "Using smart meters to estimate low-voltage losses," *IET Generation, Transmission & Distribution*, vol. 12, pp. 1206-1212, 2018.

[3] R. Billinton and R. N. Allan, *Reliability Evaluation of Engineering Systems*, 2nd ed. New York: Plenum, 1992.

[4] B. Zeng, X. Wei, D. Zhao, C. Singh, and J. Zhang, "Hybrid probabilistic-possibilistic approach for capacity credit evaluation of demand response considering both exogenous and endogenous uncertainties," *Appl. Energy*, vol. 229, pp. 186-200, 2018.

[5] Northern Powergrid, "Standard for the Application of System Monitoring," ed, 2019.