



LV Predict 2
Methodology Report
February 2026

FNC 020424-101

60376R Issue 1

Prepared for: SP Electricity North West

SYSTEMS • ENGINEERING • TECHNOLOGY

LV Predict 2

Methodology Report

Client: SP Electricity North West
Client Ref.: 0020/4300117270
Date: February 2026
Classification: COMMERCIAL IN CONFIDENCE

Project No.:	FNC 020424-101	Compiled By:	Frazer-Nash: R Augarde, M Taylor, R Raikes, and G Horne.
Document No.:	60376R		TNEI: G Edwards, R Tawn, G McFadzean.
Issue No.:	1	Verified By:	N Pready
		Approved By:	D Munn
		Signed:	

Distribution

Copy	Recipient	Organisation
pdf	Delroy Ainsworth	SP Electricity North West
pdf	Daniel Wilson	SP Electricity North West
pdf	Gordon McFadzean	TNEI
pdf	File	Frazer-Nash Consultancy

COPYRIGHT

The Copyright in this work is jointly owned by Frazer-Nash Consultancy Limited and TNEI Services Limited (together, the Joint Owners). Reproduction in whole or in part or use for tendering or manufacturing purposes is prohibited except under an agreement with, or with the prior written consent of, both Joint Owners (unless otherwise expressly agreed in writing by the Joint Owners), and then only on the condition that this notice is included in any such reproduction or use.

DISCLAIMER

This document (including any annexes, appendices and associated computer files) is issued in confidence solely for the purpose for which it is supplied. The Joint Owners accept no liability for loss or damage suffered by any party resulting from use of this document, other than as provided for by the contract under which this document has been produced.

Originating Office: FRAZER-NASH CONSULTANCY LIMITED
Narrow Quay House, Bristol, BS1 4BA
Tel: +44 (0)117 922 6242

Verification Statement

This work was carried out on behalf of SP Electricity North West (ENW) by Frazer-Nash Consultancy and TNEI under Network Innovation Allowance (NIA) project NIA_ENWL_036.

The Low Voltage (LV) Predict models and this accompanying report have been verified for technical consistency, including comparison to the LV Predict 1 Proof-of-Concept (PoC) models. As the LV Predict 2 methodology develops models beyond the PoC stage, assumptions and data sources have been further substantiated and validated to better represent the degradation of LV infrastructure. However, gaps remain in available validation data such that, although the methodology is at a level to provide some guidance on asset management decisions, it cannot yet be verified to the standard required to develop a business-as-usual software tool. The fitted parameters developed in this project should not be used in isolation from the full underlying methodology.

Executive Summary

Greater electrification of energy, including heating and transport, will increase both the level and volatility of demands across the UK's electricity network. The greatest impacts are expected on buried Low Voltage (LV) cables. It is expected that, without intervention, this changed demand on the LV system will increase the degradation rate of cables, causing higher failure rates.

ENW and other Distribution Network Operators (DNOs) must ensure the LV cable network is fit for purpose through maintenance, refurbishment, and asset replacement and reinforcement where needed. However, monitoring of performance and failure rates for buried LV cables is very limited, and DNO's asset management decisions are largely reactive; the vast majority of works on buried cables are performed after a fault with a corresponding unplanned outage to customers.

To support the expected changes in demand and to ensure the robustness of the network, DNOs will require an improved understanding of the condition of their LV cable network. This could underpin a more sophisticated risk-based approach to asset management and a shift to more proactive maintenance and replacement.

The ENW LV Predict NIA project (2021–2023), delivered by Frazer-Nash Consultancy and TNEI, conducted initial research into a modelling framework for quantifying current and future levels of underground cable and joint degradation. The scarcity of monitoring and failure datasets for these assets necessitated a combined physical and statistical modelling approach. Physical modelling sought to define explicit relationships between a cable joint's demand, its temperature, and its degradation during its lifetime of fluctuating temperatures. The statistical modelling took training data from these physical relationships to produce a robust reduced-order model able to link demand, temperature, and degradation to a large population of network assets, whose exact physical properties are not known.

This approach produced an initial Proof-of-Concept PoC demonstration of LV cable degradation modelling and showed many promising results, suggesting such a modelling framework could one day be usefully adopted by DNOs. However, it also highlighted several important areas requiring further investigation and research before the modelling framework can be used in practice.

ENW commissioned this follow-on NIA project, LV Predict 2, to further develop and extend the methodology from the first (referred to hereafter as 'LV Predict 1'). Both the physical and statistical modelling methods were revisited in this research.

The physical modelling refined the relationships between current, temperature, and degradation. A major achievement was considering the 'heat sink' effect of soil surrounding a cable to account for the lag in a cable heating up following a step change in input current. This supplies more realistic, and much less pessimistic, temperature histories for damage modelling. The link between temperature and damage was also explored further in LV Predict 2 through adding more thermally-based degradation mechanisms including chemical degradation of insulating materials. The physical model also considers older cable and joint types alongside the modern type focused on in LV Predict 1. This second project was able to leverage real examinations of cables shared by other DNOs, to provide more confidence in the validity of the methods.

The statistical modelling achieved a step-change in the size of simulation, with over 140,000 cable years simulated compared to 24,000 in LV Predict 1. These simulations are now based on topologies of real ENW networks, with a greater and more representative range of cable and joint types considered. The simulations covered a wide range

of scenarios, including variations in Low Carbon Technology (LCT) uptake and the potential changes in customer demand. The increased scale of statistical modelling was enabled by training emulators to predict network currents from load flow data, making efficient use of computing time and resources.

The complete modelling framework is now comfortably beyond the PoC stage. Implementing the LV Predict 2 model would enable informed predictions of the network and support asset-management decision-making. The Cost-Benefit Analysis (CBA) methodology reporting also produced during this project [1] illustrates the expected benefits of investing in such a targeted proactive maintenance approach, compared to random proactive intervention or the current fully reactive strategy.

This methodology report document has the following objectives:

- ▶ To present the achievements of the research enabled by the LV Predict 2 NIA project.
- ▶ To capture the development state of the methods to inform confidence in using it to support asset management decisions. Some caveats and assumptions do remain, and challenges that LV Predict 2 uncovered are documented in this reporting.
- ▶ To list recommendations for future research and explain the potential benefits. The strategy for any future work should be to triage and prioritise these recommendations to target the greatest impact on decision-making power. This prioritisation should be informed by the findings of the LV Predict 2 CBA methodology exercise [1].
- ▶ To provide a stand-alone reference for the LV Predict work to date — superseding all previous documentation on the LV Predict 1 project. As such, information from previous documentation is reproduced in this report where required.

Acknowledgements

The LV Predict 2 project team (ENW, Frazer-Nash and TNEI) acknowledge and thank Lucy Electric UK Limited, Scottish and Southern Electricity Networks (SSEN) and UK Power Networks (UKPN) for providing access to three laboratory reports conducted by RINA Tech UK Limited, of LV cables and joints [2–4] analysed as part of their SYNAPS 2 [5] NIA project. These reports of failed or failing cables informed important changes to the physical modelling in this work, and represent a positive sharing of NIA findings and experience.

Contents

1 Abbreviations	10
2 Introduction	11
2.1 Project aims	12
2.2 Modelling framework overview	12
2.3 Structure of this document	15
3 Physical model	16
3.1 Improved temperature predictions	16
3.1.1 Transient behaviour	16
3.1.2 Account of cable thermal insulation	19
3.1.3 Temperature dependence of material properties	20
3.1.4 Temperature elevation at cable joints	21
3.1.5 Phase imbalance	23
3.1.6 Soil temperature modelling	24
3.2 Updated structural FE modelling to support physical model	28
3.3 Review of damage mechanisms	30
3.3.1 Creep damage of polymer	31
3.3.2 Fatigue damage between cable and polymer joint material	34
3.3.3 Plastic damage in polymer joint material	36
3.3.4 Chemical degradation of polymer insulation	36
3.3.5 Chemical degradation of oil-soaked paper insulation	38
3.3.6 Abrasion of paper insulation	39
3.3.7 Loss of pre-load of metal electrical connector bolts leading to runaway contact resistance	40
3.3.8 Fatigue of metal electrical connector bolts	43
3.3.9 Corrosion of lead cable or joint sheaths	44
3.3.10 Fatigue failure of lead cable or joint sheaths	45
3.4 Grouping mechanisms for an updated damage model	45
3.5 Physical modelling assumptions, judgements, and limitations log	48
4 Network simulation: demand to damage	50
4.1 Sampling and simulation process	51
4.1.1 Step 1: Collating customer numbers	52
4.1.2 Step 2: Sampling smart meter, Electric Vehicle (EV) and heat pump numbers	52
4.1.3 Step 3: Sample ordering and labelling	53
4.1.4 Step 4: Cumulative customer and technology numbers	53
4.1.5 Step 5: Identifying joints and cables to model	54
4.1.6 Step 6: Soil temperatures	55
4.1.7 Step 7: Demand sampling	55
4.1.8 Step 8: Emulator	63
4.1.9 Steps 9 and 10: Physical simulation of joint temperature and damage	64
4.2 Simulation assumptions	64

5	Load flow modelling from GIS	66
5.1	Network topology from GIS data	66
5.1.1	GIS processing methodology	68
5.2	Load flow analysis	71
5.2.1	Advantages of using emulators	72
5.2.2	Emulator structure	73
5.2.3	Building emulator framework	75
5.2.4	Predicting responses	78
5.2.5	Emulator performance	78
5.2.6	Emulator assumptions	80
6	Statistical modelling of current, temperature and damage	82
6.0.1	Model target variables	82
6.0.2	Modelling of uncertainty	82
6.0.3	Out-of-sample model evaluation	84
6.1	Current models	84
6.1.1	Smart meter features	89
6.1.2	Current model evaluation	90
6.2	Temperature models	93
6.2.1	Temperature model evaluation	97
6.3	Damage model	100
6.3.1	Initial model: beta distribution	101
6.3.2	Refined model approach	103
6.3.3	Damage model evaluation	103
7	Model usage	107
7.1	Single network example	107
7.2	Whole-region demonstration	112
8	Sensitivity analyses	119
8.1	Phase imbalance sensitivity	119
8.2	EV charging behaviour sensitivity	123
9	Discussion	130
9.1	Physical models	130
9.2	Statistical models	132
9.3	Implementation considerations	134
10	Conclusions and recommendations	136
10.1	Conclusions	136
10.2	Recommendations	137
11	References	139
A	Physical model demonstrator code structure	142
A.1	Code structure	142



A.1.1	Cable specification class	142
A.1.2	Cable class	142
A.1.3	Joint class	143
A.2	Damage calculation	143
A.3	Code structure UML	145
B	Detail on technical terms	146
B.1	GAM	146
B.2	Elastic net regularisation	147
B.3	Cross-validation	147
C	Table of soil properties used in modelling	148

1 Abbreviations

AL	Active Learning
ASTM	American Society for Testing and Materials
CBA	Cost-Benefit Analysis
CET	Central England Temperature
CLNR	Customer-Led Network Revolution
DFES	Distribution Future Energy Scenarios
DNO	Distribution Network Operator
ENW	SP Electricity North West
EV	Electric Vehicle
FEA	Finite Element Analysis
GAM	Generalised Additive Model
GIS	Geographic Information System
HV	High Voltage
IPC	Insulation Piercing Connector
IPSA	Interactive Power Supply Analysis
LCT	Low Carbon Technology
LHS	Latin Hypercube Sampling
LSOA	Lower-Layer Super Output Area
LV	Low Voltage
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MC	Monte-Carlo
ML	Machine Learning
MNSTD	Mean Normalised Standard Deviation
MPAN	Meter Point Administration Number
NIA	Network Innovation Allowance
OOP	Object-Oriented Programming
PILC	Paper-Insulated Lead-Covered
PoC	Proof-of-Concept
PU	Polyurethane
R&D	Research and Development
RAM	Random Access Memory
SEN	Scottish and Southern Electricity Networks
UKPN	UK Power Networks
UML	Unified Modelling Language
XLPE	Cross-Linked PolyEthylene

2 Introduction

Distribution Network Operators (DNOs) are responsible for the maintenance of the Low Voltage (LV) networks and assets in Great Britain, ensuring the network is fit for purpose. This includes underground assets, mainly LV cables, which are subject to degradation due to various factors including environmental conditions, the magnitude and shape of the demand profiles they supply, and the number of joints between components. This degradation will eventually lead to failure and interruptions to customer supply, the avoidance of which is among the DNOs' main objectives. As energy systems evolve to meet net zero targets, it is widely expected that there will be greater electrification of energy (particularly heat and transport). This will have the effect of increasing both the level and volatility of demands across the network, with the greatest impacts expected on the LV system and in particular underground cables. Increasing demands on the LV networks are anticipated to lead to significant increases in peak demands and peak demand ramping rates in some areas, which will lead to more extreme temperature cycles for cables. It is expected that, without intervention, this increase in demand on the LV system will therefore increase the degradation rate of cables, causing higher failure rates. The effect is likely to be more pronounced for some cables due to the tendency for the adoption of Low Carbon Technology (LCT) to occur in clusters of close geographical proximity. The effect of more extreme conditions may be greater and will have a higher uncertainty on older assets, which are likely to be more degraded by decades of use.

SP Electricity North West (ENW), the DNO responsible for the operation and maintenance of the distribution network in the North West of England, plan to spend £14.3m between 2024 and 2028 to replace cables which are obsolete (over 75 years old) and deemed likely to fail, or which may pose safety risks [6]. The topic of modelling the health of underground LV cables has received relatively scant attention by researchers and innovators within industry. Current regulations and asset management methodologies typically do not cover lower voltages [7], although some aspects of the methodologies used for higher voltages could be applied to LV networks. Similarly, much of the academic research on electricity cable degradation focusses on high and medium voltage cables (1 kV and higher). This research cannot be readily applied to a low voltage network, as the primary driver for degradation in medium and high voltage cables is electrical-based degradation, which is unlikely to apply to LV cables [8].

A DNO may wish to address this issue by setting up a system of extensive monitoring of LV asset condition, given their duty of ensuring that cables are replaced before they are degraded enough to present a high risk of failure. While a widespread rollout of such an approach is being undertaken by several DNOs, this rollout is currently in progress and will take time to complete. Even when installed, monitoring only gives data on the present and recent cable usage, not historic or far future conditions, and will take time to build up a large dataset of experienced scenarios. Therefore, monitoring can be used to identify pre-faults (i.e. small defects before a fault occurs) reactively, which allows proactive fault detection but not on investment timescales. This monitoring also does not give any information about the long history of conditions the cable experienced before monitoring was installed, or allow inferences about future conditions and degradation under changing demand patterns and volumes. As a result, in most cases DNOs currently have very limited data about the usage and condition of their LV cables. To support the expected changes in demand and to ensure the robustness of the network, DNOs will require increased visibility and an improved understanding of the condition of their LV cable network. There is, therefore, a need for a modelling approach that enables a DNO to assess the state of cable degradation accurately and quickly at a particular location, using easily accessible data, and make predictions about this prior to a fault forming. Such a model is essential for the formulation of targeted and judicious programs of inspection and replacement.

Initial research into a modelling framework for quantifying current and future levels of underground cable and joint

degradation was carried out within the ENW LV Predict Network Innovation Allowance (NIA) project (2021-2023) [9], delivered by Frazer-Nash Consultancy (Frazer-Nash) and TNEI Services (TNEI). The first LV Predict NIA produced an initial conceptual demonstration of LV cable degradation modelling and showed many promising results, suggesting such a modelling framework could one day be usefully adopted by DNOs. However, it also highlighted several important areas requiring further investigation and research before the modelling framework can be used in practice. Before the LV Predict 1 project, LV network maintenance was reactive with very little consideration to proactive maintenance. Some work has been done to analyse and predict LV faults and failures based on advanced sensor measurements, but the lack of widespread adoption of these sensors means there is still a need for a method that does not depend on this data being available for the networks of interest. Consequently, ENW has commissioned this follow-on project, LV Predict 2, to further develop and extend the methodology initially developed under LV Predict 1.

2.1 Project aims

The goal of LV Predict 1 was to explore the possible structure and performance of the proposed model; this project aims to remove some of the significant simplifying assumptions made in that modelling and move from an initial demonstration to modelling of realistic network conditions. This was structured into two tasks:

- ▶ Task 1.1: Improving and validating the predictive skill and reliability of the framework
- ▶ Task 1.2: Considering how to apply the model outputs into decision making processes

Task 1.1 is achieved through considering the full range of possible topologies, cable and joint types, and physical processes contributing to cable degradation. Task 1.2 is achieved through translating the finalised modelling framework into a series of replicable steps, and reporting the practical considerations for utilising the methods in future.

This report is designed as a standalone guide to this fully developed methodology that has resulted from work in LV Predict 2, without the need to refer back to work from the initial LV Predict 1 project outputs.

2.2 Modelling framework overview

The modelling framework at the heart of the project consists of two main streams: physical and statistical modelling. This approach is needed as there is no relevant, readily available historic measured data on cable conditions and degradation, with which to explore relationships and fit purely statistical predictive models.

Physical modelling enables the simulation of a wide range of scenarios that LV networks could be exposed to, and the cable conditions these result in. Explicitly modelling physical relationships is necessary since there is no measured data on the vast majority of ENW's 20,000 km+ of buried LV mains cable. Indeed, some of the variables of interest in this project – in particular accumulated damage at LV cable joints – are not directly observable quantities and so physical modelling is the only way to gain insight into this.

At the point where the model is used to make predictions for new networks, the full network properties and high-resolution demands are not known, and so statistical models are built to predict quantities of interest (joint damage) from only the known, basic properties of the network. The statistical models are designed to be used to make predictions of network degradation for any (and potentially unmodelled) networks, such that it can be applied across the entire DNO network area. These models are trained on the outputs of the physical modelling process.

The full modelling process is shown in Figure 1 and the steps in this modelling framework are detailed over the course of Chapters 3 to 8.

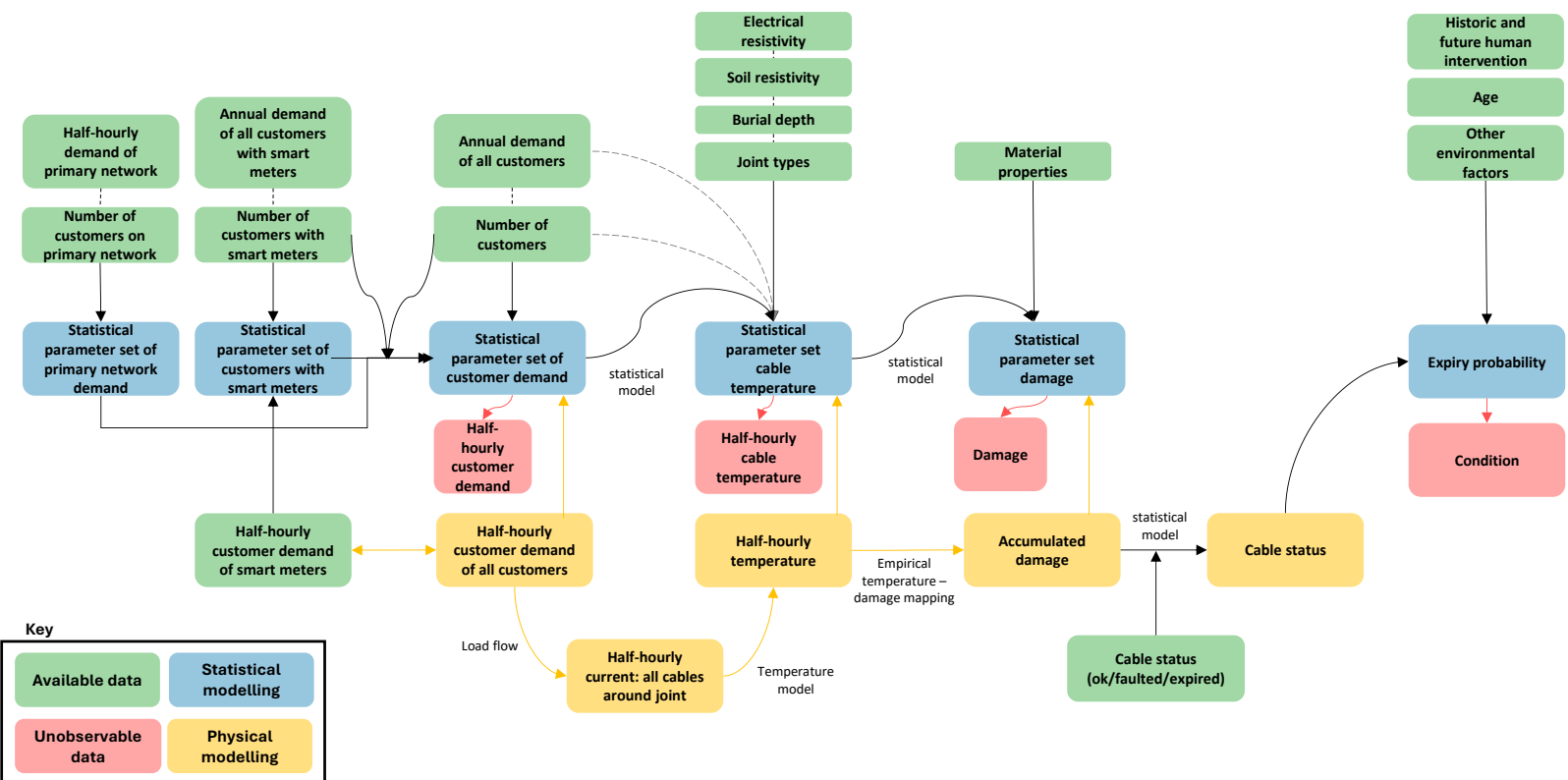


Figure 1: Schematic showing all modelling steps in the full process and how they are linked within the overall methodology.

2.3 Structure of this document

Aspects of the physical and statistical modelling methodologies are presented in several chapters of this report. This document also logs all assumptions made on each side of the modelling, as well as discussion and conclusions to inform any future work in this area.

This methodology document is structured as follows:

- ▶ Chapter 3 presents the physical modelling aspect of this project, where understanding of the heating and degradation mechanisms allows for the generation of training data for the statistical model.
- ▶ Chapter 4 presents the statistical modelling aspect of the project at a high-level. This chapter explains the complete route of simulating network conditions, with topics ranging from sampling demands to estimating corresponding damage levels in assets that see such demand. The use of the physical model to generate training data is also explained in this chapter. Chapter 4 introduces all steps in the statistical modelling process, some of which are explored in greater detail in the following dedicated chapters:
 - Chapter 5 discusses how the network topology was extracted into a usable form, and then utilised with trained emulators to predict cable currents anywhere in the network from simulated demands.
 - Chapter 6 presents the details of training a statistical model using the physical model data to allow predictions of currents, temperatures, and damage which covers many scenarios across the entirety of ENW's network.
 - Chapter 7 presents a worked example of predicting damage using the trained statistical model.
 - Chapter 8 discusses several sensitivity analyses performed to test assumptions made during modelling.
- ▶ Chapter 9 presents a discussion of the findings from both physical and statistical modelling, and the relationships between the two. This Chapter also explains practical considerations for implementing the methodology as a predictive tool
- ▶ Chapter 10 presents the conclusions on the LV Predict 2 project, and recommendations for future research or investigation.

Several annexes are included in this report to supplement the main discussion, which include:

- ▶ Annex A which details the Python programming used to demonstrate the physical modelling methods and generate training data for the statistical model.
- ▶ Annex B provides detailed explanations of some technical terms used in the methodology discussion.
- ▶ Annex C presents soil data used in temperature calculations.

3 Physical model

The previous LV Predict project [9] identified thermal degradation as the most likely leading failure mechanism for the underground LV network (alongside third-party interventions). Therefore physical modelling focused on temperature-related damage mechanisms for a cable joint with Cross-Linked PolyEthylene (XLPE)-insulated cables. A simplified damage model, fitted to Finite Element Analysis (FEA) validation data, created a Proof-of-Concept (PoC) of how to link demand to cable thermal stresses, and finally to degradation and lifetime.

In this work, the physical model was revisited to build upon that initial PoC, including a cold-eye review on assumptions in light of any new data or understanding made available in between the two projects.

The physical modelling work conducted as part of this project focused on two areas. Firstly, the temperature prediction for cables and joints has been developed. Secondly, the range of materials and damage mechanisms has been expanded based on developed understanding and research. This resulted in significant structural change to the physical damage model, with the objective of robustly predicting damage in more cable and joint types, within the computing and timescale limits of the work. However, the investigations into damage mechanisms corroborated the LV Predict 1 conclusion that temperature history, via a range of mechanisms, is the key driver of LV asset damage that should be modelled in this work.

3.1 Improved temperature predictions

The following sections discuss a series of changes made to the temperature modelling for LV Predict 2:

Transient behaviour: Capturing the time lag between a change in current and the thermal response of the cable.

Account of cable thermal insulation: Considering the insulating effects within the cable as well as the surrounding soil.

Temperature dependence of material properties: Accounting for the increase in resistivity at higher temperatures.

Temperature elevation at cable joints: Including additional ohmic heating from electrical contacts in joints.

Phase imbalance: Considering how imbalanced currents across a cable's 3 phases could incur further damage.

Soil temperature modelling: Determining accurate thermal properties for the soil in ENW's network.

3.1.1 Transient behaviour

The previous LV Predict project [9] adopted a steady-state approach to cable temperature, such that cable current could be used to calculate cable temperature so that mapping was 'one-to-one'. The sample rate was such that cable current and temperature was established for discrete steps of 30 minute duration. An inherent assumption was made such that temperature lag, due to the thermal mass of the cable and the surrounding soil, was insignificant compared with the 30 minute sample rate.

In reality, the thermal mass of the cable and the surrounding soil is significant, and steady state thermal conditions are not reached within 30 minutes following any appreciable change in current. To consider an example transient condition, a Finite Element thermal analysis was conducted to consider a 300 mm² waveform cable buried at a depth of 0.5 m and carrying a severely transient load. The modelled geometry is shown in Figure 2.

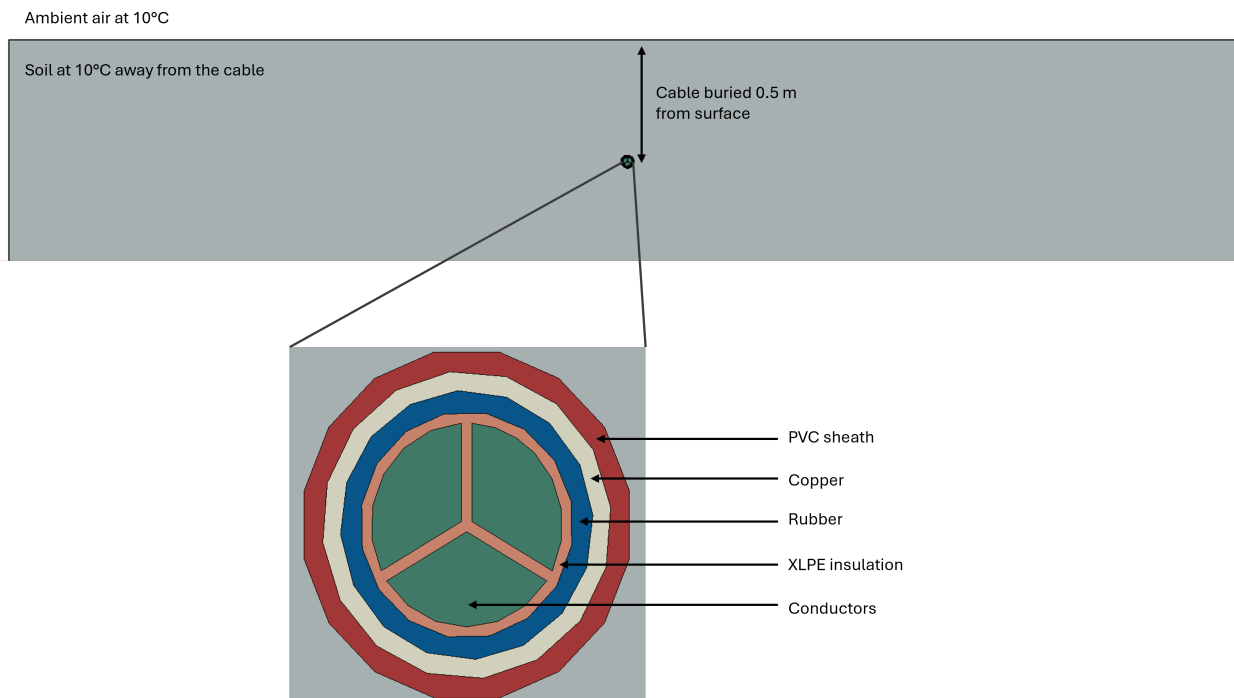


Figure 2: Geometry modelled in the transient thermal FE analysis.

The transient temperature result is compared with the equivalent steady-state temperature result in Figure 3 below. The graph on the left shows the current profile applied in each of the three conductor regions in the FEA model. The graph on the right shows the temperature extracted from one of the conductor regions in the FE model; temperature within each region was found to be approximately uniform, and comparable for all three conductor segments despite their small differences in buried depth.

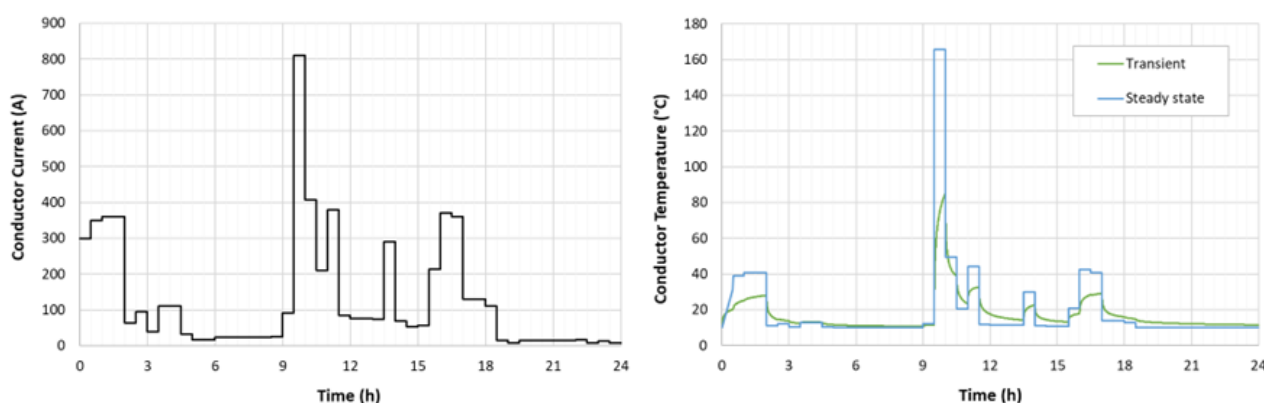


Figure 3: Comparison of steady-state and transient temperatures predicted for an example demand profile, derived from Figure 17 of LV Predict 1 Final Report [9].

The transient temperature result demonstrates that steady state conditions are not reached within the 30 minute time-step, and for severely transient conditions (e.g. the 30 minute ~800 A 'spike' in current) transient conductor

temperature varies significantly from steady state temperature. The transient result also indicates two discrete time-periods to the thermal response, which correlate with the following mechanisms;

- ▶ On step change in current, a short-duration response dominates. For example, on step change from ~100 A to ~800 A (at time = 9.5 hr), conductor temperature initially rises relatively quickly (from ~10 °C to ~60 °C). This short-duration response can be attributed to the thermal mass of the aluminium conductor section.
- ▶ On step change in current, a much longer-duration response follows the initial response. For example, on step change from ~120 A to ~20 A (at time = 18.5 hr), conductor temperature drops slowly over a period of several hours. This long-duration response can be attributed to the thermal mass of the soil surrounding the cable. Indeed, the temperature contours output by the FEA model (Figure 4) show how the bulk of the soil remains very close to ambient temperature, behaving as a heat sink. This behaviour was checked via a longer FEA analysis of a prolonged step change in current, which showed a very long decay time before temperatures reached the steady-state result.

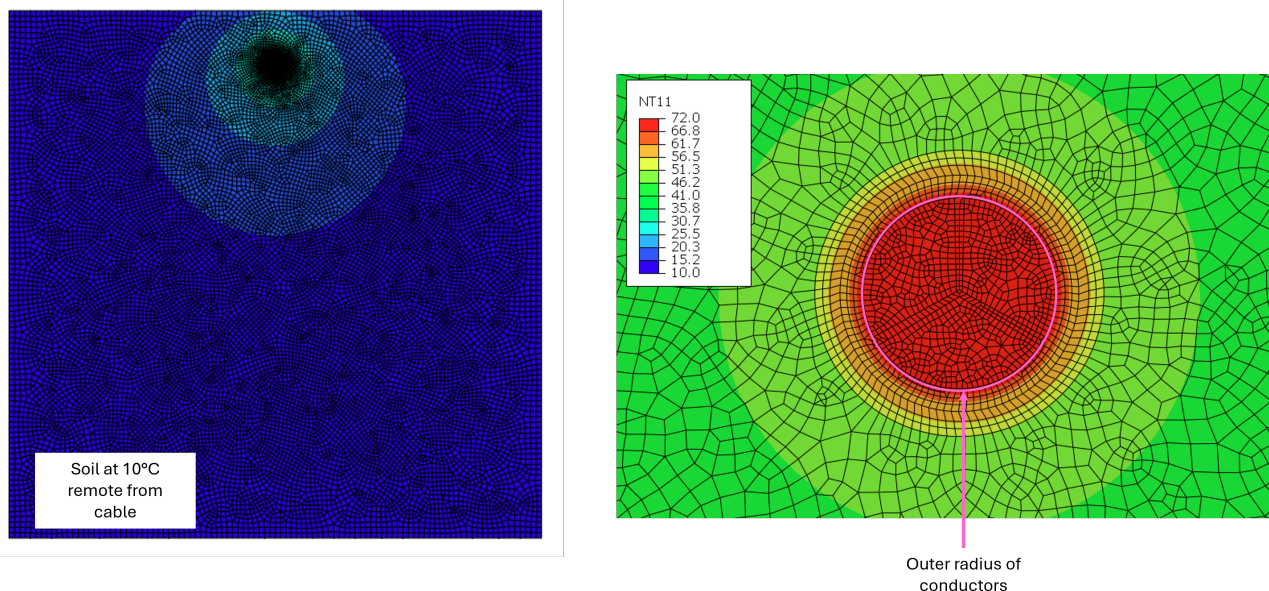


Figure 4: Temperature contours from the steady-state temperature model, showing the concentration of high temperatures around the conductor, and the heat-sink behaviour of the soil.

During development of the previous LV Predict model an approach for modelling of the thermal transient was progressed as a study. The approach applied a ‘thermal time constant’ in line with CIGRE 601 [10] to fit an exponential trend in conductor temperature following a step change in current. However it was not incorporated into the PoC as the transient results did not show significant lag against the steady-state results and therefore little benefit was demonstrated through this more computationally-expensive route. However, the review of methods in LV Predict 2 has allowed the CIGRE approach to be adapted more closely to buried LV assets — particularly to account for the thermal mass of the soil, which promotes very different behaviour to the overhead cables in air that informed the CIGRE models.

A thermal time constant, τ , was established as follows to account for the soil thermal mass:

$$\tau = n \cdot m_c \cdot c_{p,c} \cdot T_{R,c} + \rho_{\text{soil}} \cdot \pi \cdot d_c^2 \cdot f_{\text{soil}} \cdot c_{p,s} \cdot T_{R,s} \quad (1)$$

Where:

n	is the number of phases (typically 3)
m_c	is the mass of the conductor per unit length, in kg/m
$c_{p,c}$	is the specific heat capacity of the conductor material, and
$c_{p,s}$	is the specific heat capacity of the soil - both in J/kg/K
ρ_{soil}	is the density of the soil surrounding the cable, in kg/m ³
d_c	is the depth of the cable, in m
f_{soil}	is an empirical factor which crudely corresponds to the amount of soil volume which is 'tracking' the conductor temperature. A value of 0.001 was established as reasonable, by fitting to the FE study as discussed above. This value is, therefore, case specific and is not generally applicable outside of the context of this work.
$T_{R,c}$	is the thermal resistance (in Km/W) of the insulation and soil surrounding the buried cable; taken as $T_3 + T_4$ (see Section 3.1.2 below).
$T_{R,s}$	is the thermal resistance (in Km/W) of the soil surrounding the buried cable; taken as T_4 (see Section 3.1.2 below).

The time constant represents the time for temperature change of the conductor to reach $1 - 1/e$ (i.e. 63%) of the equivalent steady-state temperature change (corresponding to the step change in current). Temperature at the end of the 30 minute time step, θ , is then calculated as;

$$\theta = \theta_{SS} - (\theta_{SS} - \theta_i) \cdot \exp \frac{-t}{\tau} \quad (2)$$

Where:

θ_{SS}	is the steady-state temperature corresponding to the current running during the time step (i.e. after the step change in current).
θ_i	is the steady-state temperature corresponding to the current running prior to the time step (i.e. before the step change in current).
t	is the time step duration; i.e. 30 minutes

3.1.2 Account of cable thermal insulation

The thermal degradation modelled in LV Predict 1 [9] is driven by the cable conductor temperature, linked directly to demand through the cable, due to ohmic heating in the conductor. Conductor temperature was determined as follows;

$$\theta_{\text{conductor}} = \theta_{\text{ambient}} + n \cdot I_{\text{nominal}}^2 \cdot R_{\text{cable}} \cdot T_4 \quad (3)$$

Where:

$\theta_{\text{conductor}}$	is the calculated temperature of the cable conductor
θ_{ambient}	is the ambient soil temperature
n	is the number of phases (typically 3).
I_{nominal}	is the balanced current (in Amps) flowing in the cable (or the average current, for an unbalanced load).
R_{cable}	is the conductor resistance (in Ω/m), calculated based on conductor resistivity and conductor area.
T_4	is the 'external thermal resistance' of the soil surrounding the buried cable (in Km/W). T_4 is calculated using the method prescribed by IEC 60287-2-1 [11]

Hence, the conductor temperature is based on the thermal resistance of the soil, and no component of thermal resistance was included to account for the cable insulation. A study conducted as part of the work reported here has determined that the thermal resistance of the cable insulation is significant. For example, for a 300 mm² waveform cable buried at a depth of 500 mm in limestone backfill, the thermal resistance attributable to the cable insulation is approximately 50% of the thermal resistance attributable to the soil. Hence by not accounting for thermal resistance of the cable insulation, conductor temperature is under-estimated by approximately 30%.

To improve conductor temperature prediction, the method was updated as follows:

$$\theta_{\text{conductor}} = \theta_{\text{ambient}} + n \cdot I_{\text{nominal}}^2 \cdot R_{\text{cable}} \cdot (T_3 + T_4) \quad (4)$$

Where T_3 is introduced to account for the thermal resistance of the cable insulation. T_3 (in Km/W) is calculated using the approach provided in IEC 60287-2-1 [11], modified so that it can be applied to multiple cable types:

$$T_3 = \frac{1}{2\pi} \cdot \frac{1}{k_{\text{insulation}}} \cdot \ln \left(1 + \frac{2 \cdot t_{\text{insulation}}}{D_{\text{equiv.conductor}}} \right) \quad (5)$$

Where:

$k_{\text{insulation}}$	is the thermal conductivity of the insulation material. A constant value of 0.15 W/mK, corresponding to the typical conductivity of PVC/XLPE was adopted.
$t_{\text{insulation}}$	is the effective thickness of the cable insulation (in mm), taken as the total thickness of the insulation layers between the conductor and the cable outer.
$D_{\text{equiv.conductor}}$	is the equivalent diameter of the cable conductor bundle (in mm), taken as the cable outer diameter less two times the effective thickness of the cable insulation.

It is noted that several simplifications remain and the approach does not strictly align with that prescribed by IEC 60287-2-1 [11]. For example, methods to account for the effect of cables being buried in close proximity exist in IEC 60287-2-1 but were omitted from this model due to their complexity and the knowledge required of all surrounding cables.

3.1.3 Temperature dependence of material properties

Temperature dependence was not considered as part of the previous LV Predict Final Report [9]. The inherent assumption, that temperature dependence of material properties is insignificant, was reviewed as part of the work reported here. The assumption was judged appropriate for all properties except conductor resistivity. The increase in resistivity at elevated temperatures typical in cables under high load is significant. For example, for a 300 mm²

waveform cable running at 600 A conductor, temperature is estimated at 83.5 °C when considering a constant resistivity for the aluminium conductor (corresponding to resistivity at 20 °C). However, the temperature is estimated at 105.1 °C when considering temperature dependence of resistivity, and therefore, implemented in the model. Resistivity for copper and aluminium conductors was as shown in Figure 5.

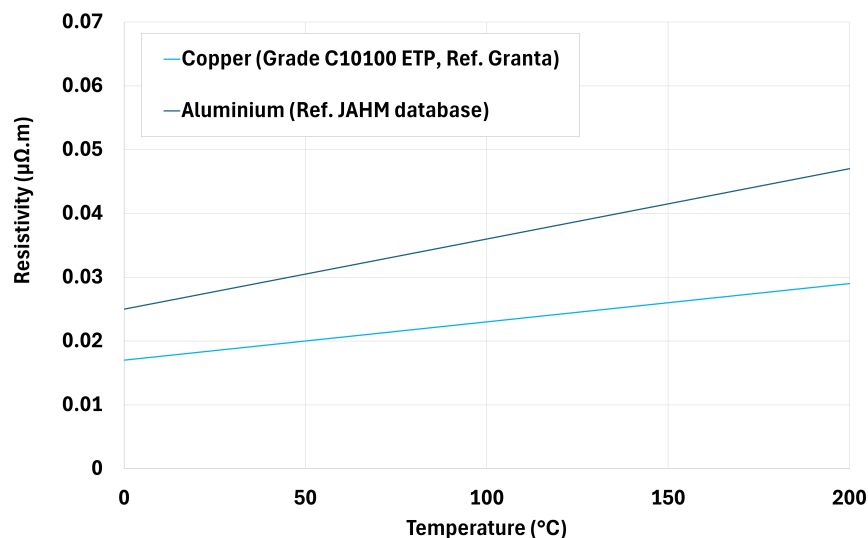


Figure 5: Variation of electrical resistivity with temperature.

3.1.4 Temperature elevation at cable joints

Local high temperature in the joint was not specifically considered as part of the previous LV Predict project [9]. As temperature elevation local to the joint can be significant, and as operational experience indicates the majority of failures in the LV network occur at joints, a method for adjustment of temperature at the joint was developed.

Three factors are considered to contribute to local temperature at a cable joint:

1. The electrical contact resistance attributable to the electrical connection will introduce a local heat source. This local heat source (on its own) will lead to a local increase in temperature.
2. The additional insulation within the joint casing will introduce a local thermal resistance. This local thermal resistance (on its own) will lead to a local increase in temperature.
3. The arrangement of the backfill (soil) local to the joint may be different to that surrounding the cable remote from the joint, in particular if the joint was retro fit. It is understood that backfill procedures for joints prescribe fine granular material, or sand, including a significant bedding depth (below the cable). The effective width of the backfilled trench is typically also greater in the region of the joint. The granular nature of the backfill, and the typically larger area of backfill in the region of the joint, will mean that the thermal conductivity of the ground is strongly dependant on groundwater conditions. When the backfill is dry, thermal conductivity will be low relative to the surrounding native soil and hence this will lead to a relative increase in joint temperature. When the backfill is wet, thermal conductivity will be high relative to the surrounding native soil and this will lead to a relative reduction in joint temperature.

The model has been updated to address the first two points above, forming the two terms in Equation 6. No updates have been made to consider the effects of the local backfill arrangements, and this represents a simplification in the model; the simplification is judged appropriate on the basis that credible variation in thermal conductivity of the backfill is expected to result in reasonably small variation in cable temperature.

Test studies applying FEA were conducted to understand the effects of electrical contact resistance and local insulation at the joint. The studies showed that thermal conduction along the length of the cable conductor is significant, i.e. the solution requires a 3D understanding of heat flux (heat flux along the conductor as well as heat flux via the soil in the plane perpendicular to the conductor). Inclusion of a 3D model was judged too complex for implementation in the LV Predict model and therefore a simplified expression for local temperature elevation at the joint was developed.

The first term in Equation 6 is a function of the electrical contact resistance, causing ohmic heating as seen in the cables themselves. The amount of heating is scaled with geometric and material parameters for the joint and cable discussed below. The second term in Equation 6 accounts for the higher temperature experienced by the cable encased in the polyurethane joint insulation, which is both thicker and less thermally conductive than the cable's own insulation. These two differences are accounted for in Equation 6 with a factor of 5 on the thermal resistance component T_3 , normalised against the cable-only thermal resistance (T_3 itself) and geometric and material factors of the joint and cable:

$$\Delta\theta_{\text{joint}} = n_c \cdot I_{\text{nominal}}^2 \cdot \frac{R_{\text{contact}}}{2} \cdot \frac{l_{\text{char}}}{k_c \cdot A_{\text{cable}}} + n_c \cdot I_{\text{nominal}}^2 \cdot \frac{R_{\text{cable}}}{2} \cdot \frac{l_{\text{char}}^2}{k_c \cdot A_{\text{cable}}} \cdot \frac{T_3}{5T_3} \quad (6)$$

Where:

$\Delta\theta_{\text{joint}}$	is the local temperature increase at the joint; i.e. a temperature to be added to the conductor temperature calculated for the 'plain' cable, as described above in Section 3.1.2.
n_c	is the number of bolted electrical connections made; e.g. 3 for a splice joint in a 3 phase cable.
I_{nominal}^2	is the balanced current in Amps flowing across the electrical connection; e.g. the 'main' cable current for a splice joint.
R_{contact}	is the electrical contact resistance attributable to a single contact (i.e. a single phase). Contact resistance is controlled by design of the electrical connector (e.g. shear bolts to control contact load) and jointing procedures (e.g. for surface impurities), however significant variation is expected. A range from 1-10 $\mu\Omega$ is judged credible for modern connections, based on calculation to the method provided in PD IEC/TR 60943 [12], which has been used to predict increase in contact resistance over time (see Section 3.3.7)
l_{char}	is a characteristic length. An empirical relationship based on joint shell length and cable area was found to provide a reasonable match to finite element results obtained for test cases; $l_{\text{char}} = l_{\text{shell}} + A_{\text{cable}}/2$ (where lengths are in mm and A_{cable} is in mm ²).
k_c	is the thermal conductivity of the conductor.
A_{cable}	is a the cable area (cross sectional area per phase).
n	is the number of phases (typically 3).
R_{cable}	is the cable resistance, calculated based on temperature-dependent conductor resistivity and conductor area of the largest cable going into the joint (by A_{cable}), as this drives the length of the joint shell, governing the characteristic length l_{char} .

3.1.5 Phase imbalance

Phase imbalance on the LV network is understood to be significant; exploratory study conducted as part of this work has determined a median demand phase imbalance ratio of around 1.25 (see Figure 90 in Section 3.1.5). Note that phase imbalance ratio is defined as the maximum phase current divided by the mean current, and is hence analogous to the definition of phase imbalance provided by NEMA MG 1 [13].

The effects of phase imbalance are not included in the LV Predict model, as it has not been possible to develop a reliable understanding of the corresponding damage mechanisms (and hence no benefit is achieved by introducing phase imbalance as a variable). An inherent assumption is hence made in the LV Predict modelling+ that phase currents are always balanced. To understand the significance of this assumption with regards cable and joint temperatures, the following notes present the findings of an exploratory study.

Phase imbalance results in a temperature difference across parallel conductors (phases), due to the difference in ohmic heating of each phase. The temperature difference will depend on the difference in ohmic heating and on the heat flux between conductors (phases). Within the packaged cable, heat flux between conductors will be significant due to their close proximity, and therefore temperature difference will be relatively small. At the joint, the conductors are splayed to facilitate jointing and an insulating material is used to fill the void between conductors (expanded polyurethane is used on modern joints); hence heat flux between conductors will be small and therefore temperature difference will be relatively large.

To investigate the scale of the temperature difference, two scenarios were considered to estimate temperature difference within the packaged cable and within the joint shell. To consider the packaged cable, a 2D finite element model of a buried waveform cable was used to estimate the temperature difference across conductors (see Table 1 below). To consider the joint, the method presented in Section 3.1.2 was used to estimate the temperature for levels of phase current corresponding to a prescribed phase imbalance ratio (see Table 2 below).

	Cross-Sectional Area per Conductor	
Phase imbalance ratio	95 mm ²	300 mm ²
1.02	0.2 °C	0.2 °C
1.25	4.4 °C	3.7 °C

Table 1: Temperature difference within the packaged cable; estimated via Finite Element analysis.

	Cross-Sectional Area per Conductor	
Phase imbalance ratio	95 mm ²	300 mm ²
1.02	4.6 °C	4.6 °C
1.25	62.2 °C	62.2 °C

Table 2: Temperature difference at the joint; estimated via calculation per Section 3.1.2.

Within the cable, temperature difference is relatively small (Table 1). For credible levels of phase imbalance, mechanical loads (due to constrained thermal expansion) are expected to be small enough such that associated damage is negligible.

At the joint, temperature difference is significant. Further, geometry effects (in particular, axial load will induce bending stress in the splayed conductor) and stress concentrations (at the electrical connector) will result in stress

enhancement. Based on initial simplified hand calculations, stresses within the joint beyond 100 MPa are judged likely for phase imbalance reaching 25 %. Hence significant damage due to phase imbalance is credible. While this could not be considered in the timescales and scope of this project, it could be valuable area of further research.

3.1.6 Soil temperature modelling

While the temperature within the cables and joints is driven by the heating effects caused by current flow, the ability of this heat to dissipate into the surrounding soil also affects the resulting temperature the cables and joints experience. There are two soil-dependent inputs needed for the cable thermal modelling: the relevant thermal properties of the soil surrounding the cable, and a time series of soil temperature (θ_{ambient} in Equation 4).

3.1.6.1 Soil thermal properties

The soil thermal resistivity and cable burial depth are required for cable temperature calculations. Soil thermal resistivity is extracted for the cable's location based on a shapefile map of soil types and a mapping of soil type to thermal resistivity (as shown in Figure 6). Soil thermal resistivity varies by moisture content and typically dry and wet resistivities are stated for each soil type. The overall thermal resistivity is then a weighted average of these, based on mean rainfall for that location, calculated through the steps:

1. Map the different soil types across the ENW operating region, as shown in Figure 6a. A similar mapping is produced for average rainfall by location. Combining these with the set of simulated networks gives a soil type and average rainfall per location.
2. Dry and wet thermal resistivities [14, 15], specific heat capacities [16, 17], and densities [17] are merged in by matching on soil type. Where soil types are mixed, a weighted average of the soil constituents is taken, based on their relative proportions.
3. The average rainfall, r , was used to weight the wet (x_{wet}) and dry (x_{dry}) values of soil properties to get an overall value x_{avg} for each location, according to the formula

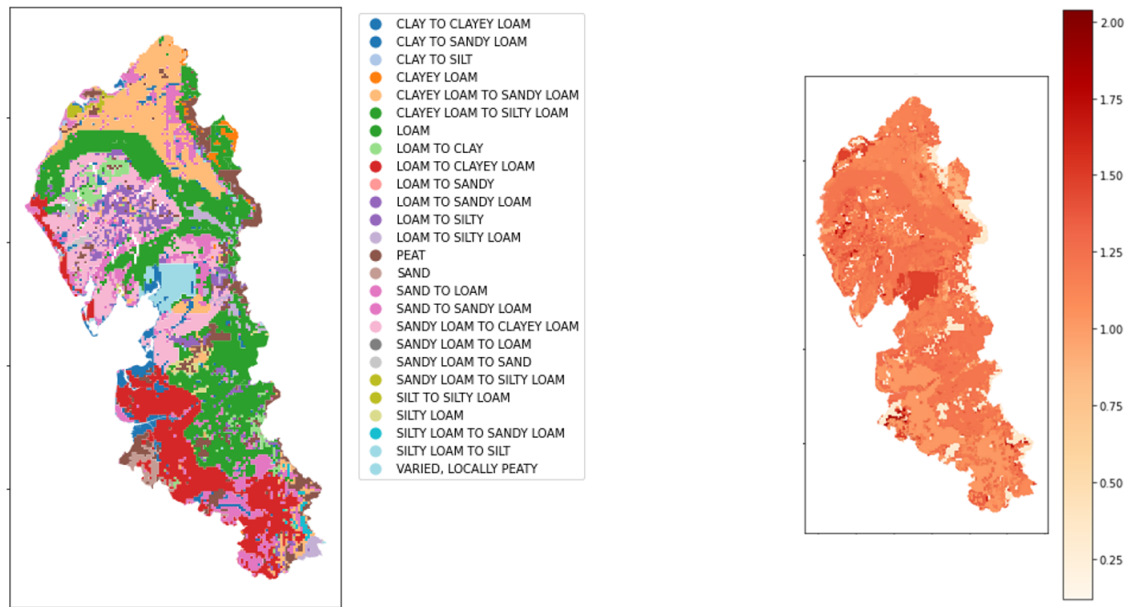
$$x_{\text{avg}} = x_{\text{dry}} + (x_{\text{wet}} - x_{\text{dry}}) \times \frac{\min(r, 10)}{10} \quad (7)$$

This process for calculating rainfall weighted properties was applied to thermal resistivity, specific heat capacity and soil density. The factor $\min(r, 10)/10$ accounts for the fact the soil is completely saturated for rainfall values over 10 mm/day.

This method could be further refined to calculate time varying soil properties that change dynamically with weather and assumed drainage rates, rather than assuming static values based on average rainfall for the location.

The table of wet and dry values used is supplied in Appendix C. All cables are assumed buried at a 0.5 m depth, as was assumed in the FE presented in Chapter 2.¹

¹ Confirmed by D. Ainsworth as part of LV predict phase 1 workshop on cable and joint failure, 03/11/2021.



a: Map of soil types across the ENW operating region.

b: Average soil thermal resistivity across the ENW operating region.

Figure 6: Soil properties across the ENW region.

3.1.6.2 Soil temperature modelling

In addition to average soil properties for every location, a time series of soil temperatures at the joints must be supplied to the physical model of temperature. This temperature data is supplied for the same time points as the cable current time series. Soil temperature data at the required locations (and realistic cable burial depths) are not available, as such, the development of a model to predict soil temperature for any given location was required.

Ground temperature is less variable, and therefore more predictable with much lower uncertainty, than air temperatures, especially at the 0.5 m depth typical for underground LV cables. As such, a deterministic model of soil temperature is built, and uncertainty on soil temperature is ignored (although the model remains statistical in nature). A Generalised Additive Model (GAM) was chosen for this model due to its ability to represent smooth relationships, but still in an additive framework where the form of fitted individual effects can be examined. The model was trained on the MIDAS dataset [18], an open dataset of meteorological data collected at UK Met Office monitoring stations, that includes daily soil temperature at depths of 0.01–0.1 m. This ground data was supplemented with records of nationwide air temperature from the UK MET Office Hadley centre observations database [19]. Central England Temperature (CET) is calculated over a triangular area enclosed by Lancashire, London and Bristol. While this does not cover the whole of the UK, it is available in daily resolution. Monthly resolution UK-wide temperature measurements are also incorporated, using a rolling mean to smooth before averaging with the daily CET measurements, to get an overall average air temperature.

Since there is a lag between air temperature and ground temperature which increases with depth, a rolling mean temperature representing longer-term memory of recent conditions was also created. The length of window used was informed by the average distance between the averaged air temperature and the soil time series, with the final value of 15 days chosen to maximise the correlation coefficient of the model predictions. Figure 7 shows an example

time series of this rolling mean air temperature feature and the corresponding soil temperatures which follow the same oscillations. Note soil temperature is modelled at daily resolution (as opposed to the half-hourly resolution of cable demands and currents) due to data availability and the inherently slow variations seen in soil temperature at a depth of 0.5 m. Therefore the same soil temperature value is used for all time points in the same day.

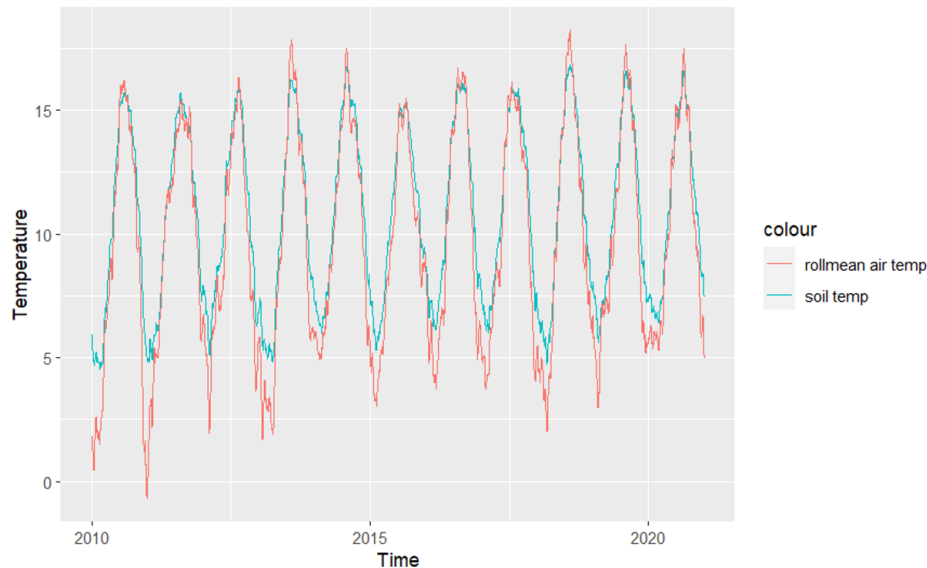


Figure 7: Soil temperature and averaged air temperature time series.

Table 3 shows the variables included as inputs for the soil temperature model and the notation used to represent them in the model equation. The rural or urban classification indicator is taken from Government data for Lower-Layer Super Output Areas (LSOAs) in England and Wales [20], and the Scottish equivalent, 'Datazones' [21].

Feature	Symbol
Soil temperature (target), °C	T_s
Day of year	d_y
Latitude, °	l
Elevation, m	e
Burial depth, m	d
Average daily air temperature, °C	t_d
Rolling average (smoothed and lagged) air temperature, °C	$\bar{t}$
Rural/urban classification indicator	R

Table 3: Soil temperature model features.

The soil temperature model includes both linear and smooth nonlinear relationships and interaction terms, as specified in Equation 8: $s(\cdot)$ are smooth, nonlinear spline terms.

$$T_s = s(d_y) + l + l^2 + s(l, d_y) + s(e) + s(d) + s(e, d) + s(d, d_y) + s(t_d) + s(t_d, d) + s(\bar{t}) + s(\bar{t}, d) + R \quad (8)$$

While this is a complex structure with several features and their interactions, the marginal effect of each can be

understood through plots. Examples of the effects of single-feature splines (the smooth nonlinear functions) are shown in Figure 8, illustrating the seasonal form of day of year, where day 0 and day 365 'join up', and the positive relationship between the engineered rolling air temperature feature, and soil temperature. The bivariate spline of depth and day of year is shown in Figure 9, with the cyclical form for day of year in the vertical direction and the lagged effect of air temperature cycles at increasing depths (the diagonal stripe of higher marginal values).

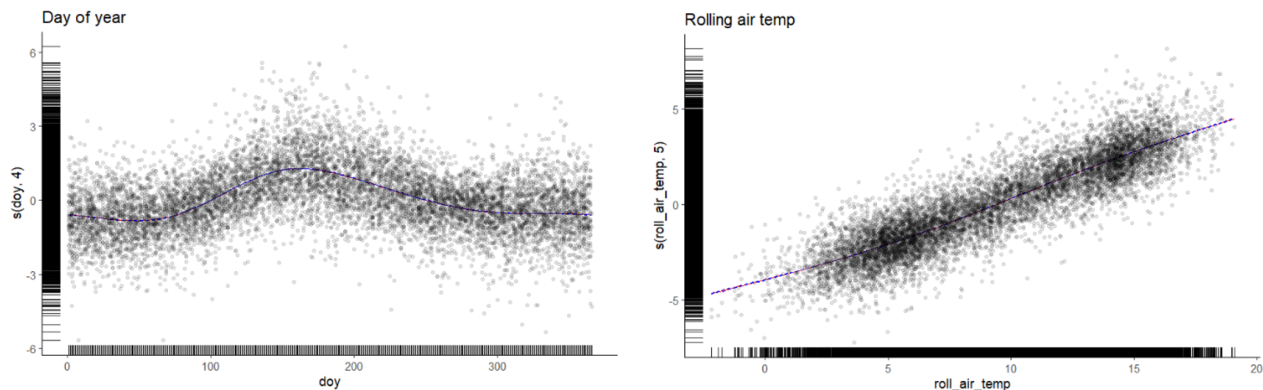


Figure 8: Examples of the marginal effects of soil temperature model features. The left plot shows the variation of ground temperature per day of year, and the right plot shows the relationship between rolling average air temperature and the ground temperature. The purple lines show the average values. Temperatures are in °C

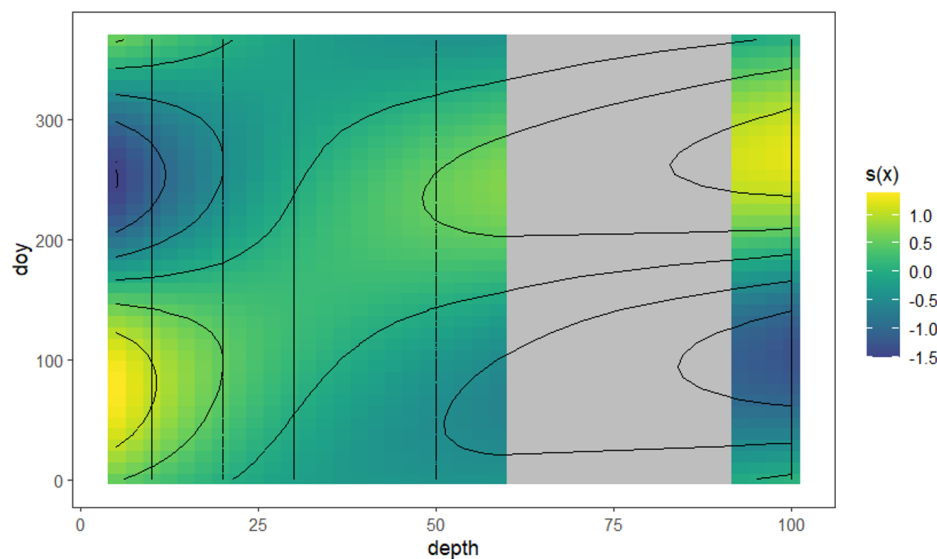


Figure 9: Bivariate spline of depth (in centimetres) and day of year in soil temperature model, where the grey region lacked input data and has been extrapolated across. Colour shows the *marginal* additive effect of this term in the model, in °C

While for this project the burial depth of interest is 0.5 m, including soil temperature measurements at other depths (with a model feature accounting for depth) increased the size of the training dataset, thereby improving the model fit. It created a generalisable model that could be used to predict other soil depths if needed. The use of cross validation on the training dataset meant that predictions could be made for every location with soil temperature measurements, and their accuracy calculated. It was found that the fitted soil temperature model has good

predictive power, with an R^2 value of 0.966 and Mean Absolute Error (MAE) of 0.67 °C at a depth of 0.5m.

This means that the model can correctly predict the variability in soil temperature across a large set of location and weather combinations. This also gives confidence in predicting soil temperatures across other depths.

Outputs of this model are then used as inputs for the physical cable temperature modelling, which provides the training data for the statistical cable temperature model.

3.2 Updated structural FE modelling to support physical model

The physical modelling work in LV Predict 1 [9] applied conductor temperatures, based upon demand, to a simplified FEA model of an LV cable joint. The thermal stresses output from this FEA were used to fit a joint stress calculation, to predict the joint stress for a given increase in conductor temperature. This simple joint stress calculation could then be deployed to generate a continuous time series of stress directly from temperature data. A joint damage model then used this stress time series to calculate contributions from fatigue crack growth, creep damage, and plastic damage, and sum them to predict a joint lifetime. These calculations formed the physical model, which took example datasets of half-hourly customer demand, and output corresponding training datasets which linked that demand directly to damage accumulated.

The physical model and its supporting stress FEA logged several assumptions in order to demonstrate PoC within the project constraints.

The FEA reported in Section 4.1 of LV Predict 1 Final Report [9] used an axisymmetric model of the cable, with a postulated large area of unbonded joint filler at the centre of the cable joint. Stresses were induced by fixing the conductor temperature to create a radial temperature gradient through the cable and joint layers and the surrounding soil. The stress at the edge of the unbonded filler region was used in damage calculations.

This work produced an updated stress FEA model. Its geometry is presented in 10. The axisymmetry is retained, with thermal loading still produced from a fixed conductor temperature. The unbonded area growing from the centre of the joint is removed; the assumption of a central crack location was reviewed against experience from the ENW Training Academy and observations in the SYNAPS 2 NIA project [5]. Laboratory investigation reports from SYNAPS 2 [2–4], refer to detailed reports of cables and joints which failed, or were close to failure, before being excavated and dissected to investigate the state of different materials. These cables and joints are assets of other DNOs (UK Power Networks (UKPN), Scottish and Southern Electricity Networks (SSEN)).

It was judged that a more realistic location for crack initiation and growth could be at the end of a joint, in particular at the interface between joint filler, cable, and soil. From this site, fatigue cracks could grow inwards to the joint. Cracks originating at the outside of the joint would be susceptible to water ingress, which in turn could also accelerate crack growth. This degradation would eventually create a route for water to the joint's conductors, causing a fault. While cracks growing out from the joint would also eventually create this path for water, the growth would not be exacerbated by continuous water ingress to the advancing crack. Therefore it is judged that cracks growing inwards are likely to be a greater threat to joints than those growing outwards.

As such the updated FEA model uses a defect-free geometry, with the aim of identifying areas of stress concentration where defects are likely to initiate throughout life. This removes the need for an initial assumption about the disbond location. This is discussed further in Section 3.3.2 which covers fatigue damage.

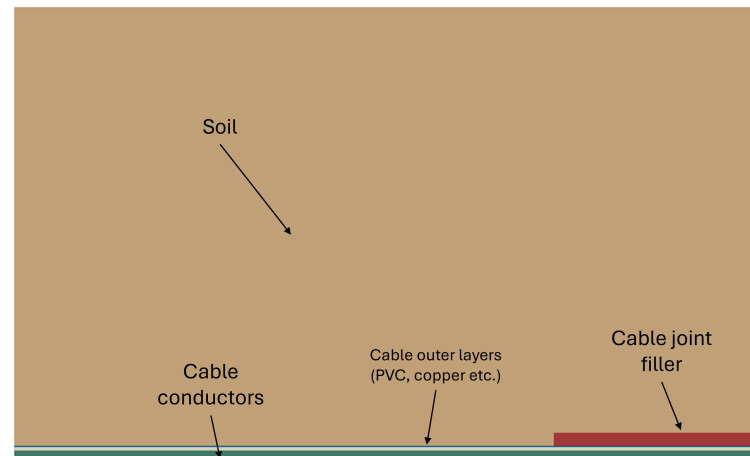


Figure 10: Axisymmetric model of a cable joint and the surrounding soil, with the three conductors approximated as a single conductor (green).

The geometry above was produced for 95 mm² and 300 mm² three-phase cables (with these conductor areas approximated as a single conductor in this axisymmetric analysis). Some specific inputs were updated, based on newly available data. For example, the stiffness of the Polyurethane (PU) joint resin was increased compared to the analysis in LV Predict 1 [9]. Initially set at 30 MPa, in line with the typical required value in ENW's specification [22], it was revised upwards to 100 MPa to align with the typical ranges for PU in wider literature [23]. This also accounts for the fact that the 30 MPa value is listed in Electricity Specification 400R10 [22] as the 1% secant modulus, indicating this value is an underestimate of stiffness at low strains.

The output of these stress FEA models indicates the locations of stresses relevant to the damage mechanisms considered in the physical damage model, where stress values can be linked to the input conductor temperatures. In this work, this is used to create simple expressions for stresses based on temperatures, which form the link between the LV Predict 2 temperature and damage modelling. The stresses extracted are discussed for each damage mechanism in the next section.

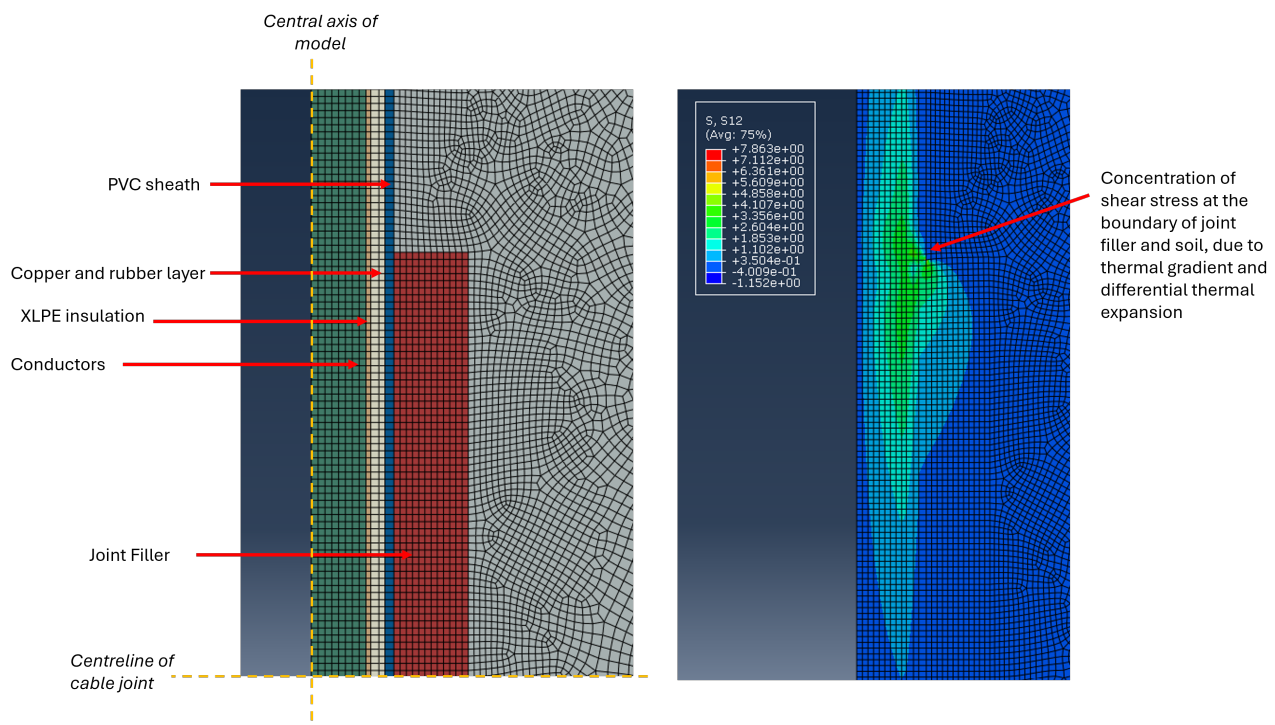


Figure 11: Axisymmetric model of a cable joint and the surrounding soil, with the three conductors approximated as a single conductor (green).

3.3 Review of damage mechanisms

The assumptions of the damage calculations in the physical model of LV Predict 1 have also been reviewed for this work. For example, only three degradation mechanisms were considered in LV Predict 1, with other modes such as electrical damage and chemical degradation discounted based on literature reviews and/or sensitivity studies. The LV Predict 1 PoC physical model was also limited to Wavecon XLPE-insulated cables only.

Research on failures in the older Paper-Insulated Lead-Covered (PILC) cables is now available [2–4], and illustrates the diversity of failure modes in these older assets, alongside the corresponding older bitumen-sealed joints. Further operational experience of joint materials and assembly was gained through a visit to ENW's Training Academy in Blackburn, which demonstrated the range of geometric details introduced, from connecting service cables to each phase of a conductor, to sealing a joint housing to prevent water ingress. The demonstrations and discussion also highlighted site specific challenges including controls for moisture and grease.

With a greater understanding of how cables and joints are assembled and how they fail, it was concluded that a significant number of failure mechanisms could be considered in the physical model, beyond the three initially treated in LV Predict 1 [9]. Therefore the work on updating the physical model damage metrics began with a list of possible further mechanisms to investigate, as illustrated in Figure 12.

An updated physical model could treat these mechanisms directly, as had been done for the existing mechanisms — or there was the option of greater structural changes to the methods to group mechanisms together at a higher level. For example, all physical processes governed by dwells at high temperature could be grouped together by a single temperature-damage relationship. Both options are limited by the lack of experimental data and many

sources of uncertainty (joint construction, material properties, etc.), but isolating mechanisms allows for using wider literature to substantiate formulae and assumptions. A simplified grouping of mechanisms has little to no supporting evidence, mainly due to the very small number of failed joints which have been examined, and the scarcity of lifetime temperature histories for those that were examined.

Examining how to model each of the new mechanisms in turn also allowed for the credibility of each failure type to be evaluated. For example, by sourcing values like activation energies or endurance limits, some of the postulated mechanisms were found to not be relevant for the temperatures and timescales in this work. Therefore, some of the mechanisms suggested at the qualitative stage have been excluded during the process of quantifying them for the physical model.

The sections below detail the extended list of damage mechanisms, how they were investigated, and which mechanisms present robust methods that can be included in an updated physical model. Each quantified mechanism's damage total is driven by time at elevated temperatures, or by the nature of thermal cycles. Where mechanisms are ruled out, the assembled evidence and arguments are presented, with suggestions on what further work or data would be needed to robustly model them in future.

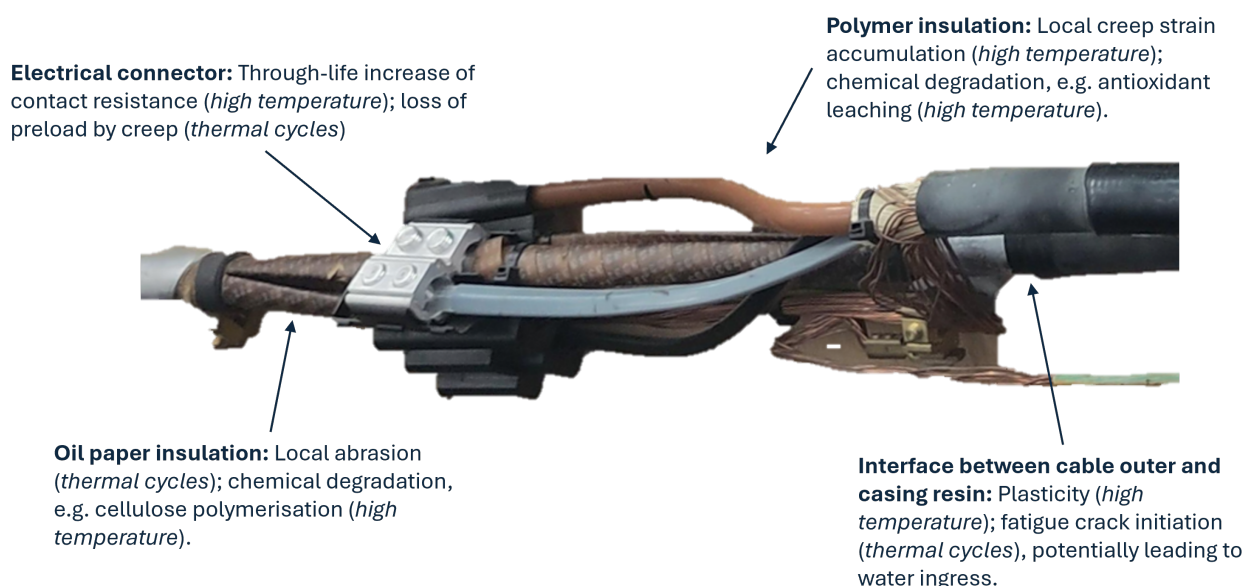


Figure 12: Identified damage mechanisms and their locations. This example joint, photographed at ENW's Training Academy, connects XLPE-insulated conductors to older paper-insulated conductors.

3.3.1 Creep damage of polymer

The polymer insulation of conductors in modern waveform cables will sustain creep damage during a lifetime at elevated temperatures as creep strain is accumulated in the material. The polymer filler in corresponding modern joints will also creep strain when operated at high temperature. As discussed in Section 4.2.2 of LV Predict 1 Final Report [9], creep strain can be assumed to accumulate via stress relaxation in the polymer during each temperature dwell. The rate of creep strain accumulation is driven by temperature. For the polymer joint filler, strain accumulation is explicitly calculated in the PoC model of LV Predict 1 [9]. This assumption has been retained and therefore the relationships between stress and creep strain accumulation are unchanged. These are detailed below for reference.

Note however, that the temperatures input to the methods below are refined by the temperature modelling work described above, and by the use of updated FEA to predict the stress.

3.3.1.1 Stress estimate for creep damage

The stress in the polymer joint filler and nearby polymer cable is derived from the FE models discussed in Section 3.2.

For the creep damage, the relevant stress was chosen as the maximum principal stress in the filler material. Figure 13 shows the stress contours of the filler material (with the rest of the geometry hidden), alongside a graph of these stress values downwards away from the top-left corner, where the stress is discontinuous. Near to this discontinuity, the stresses are very sensitive to the chosen FE mesh size, meaning a certain level of mesh refinement and distance from the discontinuity must be used to ensure physically accurate stresses are extracted. The line plot in Figure 13 includes results for two different mesh sizes to illustrate that the predicted stresses become independent of the mesh size at around 6 mm from the stress discontinuity. Hence these are the stresses extracted for use.

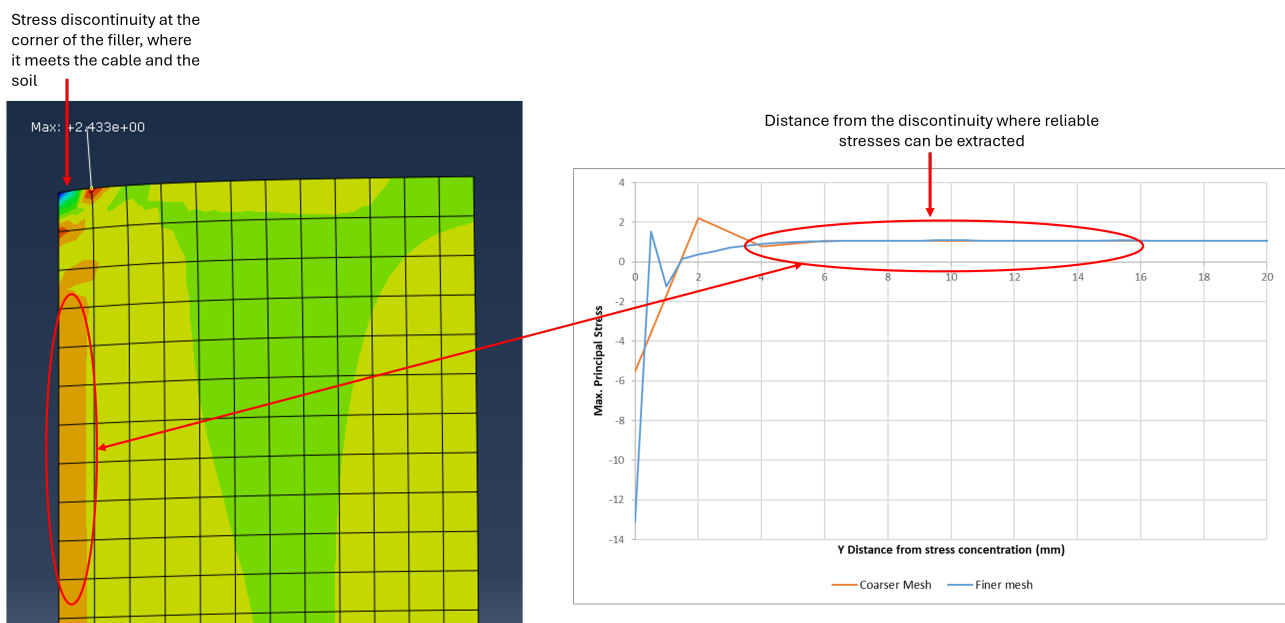


Figure 13: Extract of the FE model, showing how the maximum principal stresses were extracted for use in the creep model, with consideration of the sensitivity of the results to the FE mesh used.

A simple empirical expression was written based on the FE output, linking a joint stress value to the input conductor temperature:

$$\text{Stress for creep in joint (MPa)} = \frac{\text{Conductor Temperature} - \text{Soil Temperature}}{86.18} \quad (9)$$

The time series of stress is then interpreted into both dwell and cycle information, which are used by the creep and fatigue (see the next section) respectively.

3.3.1.2 Creep strain accumulation

The creep model uses a forward primary creep law, as presented in Equations 4.3 and 4.4 of Betten [24] with the material experiencing time-hardening over a constant stress creep dwell time t :

$$\dot{\epsilon}^c = C t^m \sigma^n \quad (10)$$

To track the relaxation of stress and the accumulation of creep strain, the above can be rearranged to give the below, relating the stresses at the start (σ_0) and end (σ) of the creep dwell of duration t :

$$\sigma = \left\{ \sigma_0^{(n-1)} + \frac{1}{Z} \left(\frac{n-1}{m+1} \right) C E t^{m+1} \right\}^{\frac{-1}{n-1}} \quad (11)$$

The coefficient C is temperature dependent. Experimental data often reports C values at specific test temperatures. The Arrhenius Law (discussed further in Section 3.3.4) links C to a temperature-independent constant A , such that the value of C at any temperature can be calculated. The $C(T_{\text{ref}})$ is also scaled with the relative yield strength of the material from Cui et al [25] compared to the known yield strength of the joint polymer in scope of this work.

$$C(T_{\text{ref}}) = A * \exp \frac{-Q}{RT_{\text{ref}}} \quad (12)$$

$$C(T) = C(T_{\text{ref}}) * \exp \left\{ \frac{Q}{R} \left(\frac{1}{T_{\text{ref}}} - \frac{1}{T} \right) \right\} \quad (13)$$

Creep damage is taken at the ratio of creep strain to maximum creep strain (usually termed 'creep ductility'). The material properties used for this damage mechanism are tabulated below, taken from Cui et al [25].

Symbol	Value	Description
$C(T_{\text{ref}})$	0.000008	Fitting Parameter at reference temperature
T_{ref}	20 °C	Reference temperature for C parameter
m	-0.75	Creep strain rate time exponent
n	1.5	Creep strain rate stress exponent
Q	7.416	Activation Energy (kJ/mol)
Z	3	Elastic Follow-up Factor
R	8.31	Molar Gas Constant (8.31J/K/mol)
ϵ_{max}^c	0.10	Maximum allowable creep strain

Table 4: Creep properties assumed for joint and cable polymer.

3.3.2 Fatigue damage between cable and polymer joint material

Defects in cable materials, or disbonds between materials, will initiate and grow under cyclic thermal loading. The FE analyses in the previous project highlighted that a potential high risk area for fatigue crack growth is between cable insulation and the joint filler material (assumed to be polymer insulation and resin filler, respectively). The crack was postulated to grow outwards from the centre of the joint to the edge, at which point the crack would provide a complete route for water ingress. Limited experimental data was available for the polyurethane, such that fatigue crack growth data from an epoxy polymer with a similar material response was utilised.

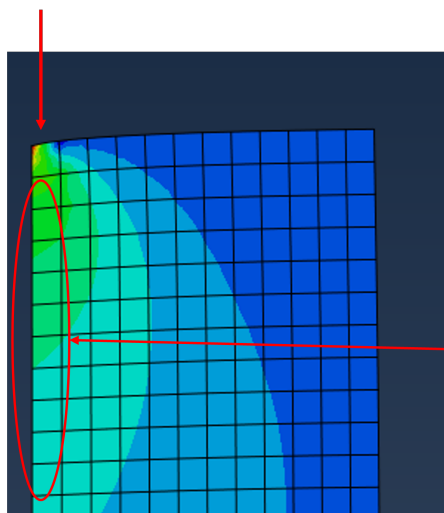
As discussed in Section 3.2, the interface between joint filler, cable, and soil is the area of interest for fatigue damage. The results indeed highlight a stress concentration in this location, indicating it is a likely site of fatigue crack initiation.

Based on new experimental data, the focus of the fatigue damage calculation is changed from crack growth to crack initiation, discussed further in Section 3.3.2.2.

3.3.2.1 Stress estimate for fatigue crack initiation

In-plane shear stress is the relevant stress measure for the fatigue damage calculation, as this corresponds to the most credible mode of defect growth between joint filler and the cable (mode II). As for the creep model, this stress is extracted from the location of interest in the FE model, illustrated below in Figure 14.

Stress discontinuity at the corner of the filler, where it meets the cable and the soil



Distance from the discontinuity where reliable stresses can be extracted

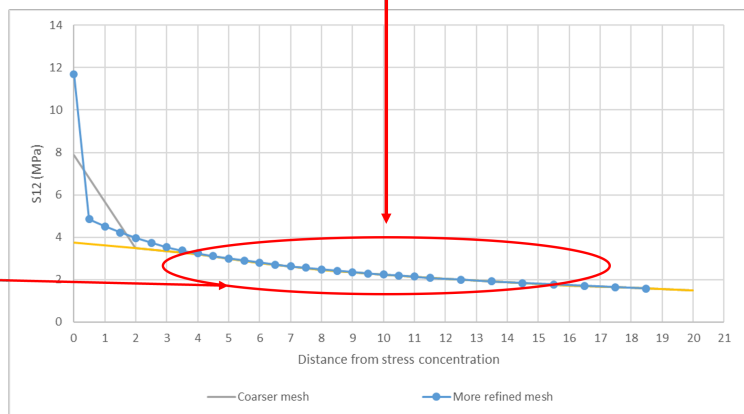


Figure 14: Extract of the FE model, showing how the maximum principal stresses were extracted for use in the creep model, with consideration of the sensitivity of the results to the FE mesh used.

A similar simple expression is derived to link the temperatures to a fatigue stress to use in damage calculations:

$$\text{Stress for creep in joint (MPa)} = \frac{\text{Conductor Temperature} - \text{Soil Temperature}}{28.05} \quad (14)$$

Stress cycles are extracted from the output of Equation 14 using the rainflow approach, as described in Section

4.2.1 of the LV Predict 1 Final Report [9]:

The rainflow approach is chosen as it the most common cycle extraction method, and there are a large range of published software libraries that can be used to implement this algorithm. The rainflow method was implemented using the Python “rainflow” package [26], which is based on the American Society for Testing and Materials (American Society for Testing and Materials (ASTM)) E1049-85 standard rainflow cycle counting algorithm [27].

Stress dwells are extracted from continuous stress data by discretising the stress time series and recording the total duration that the stress remains at a specific stress level.

3.3.2.2 Fatigue damage initiation calculation

Updated fatigue data for the joint filler at this interface has been located since the previous work. Experimental data is presented in the literature [28] for a PU adhesive joint under cyclic shear stresses at a range of temperatures. S-N curves are available for -10 °C, 20 °C, 50 °C and 80 °C.

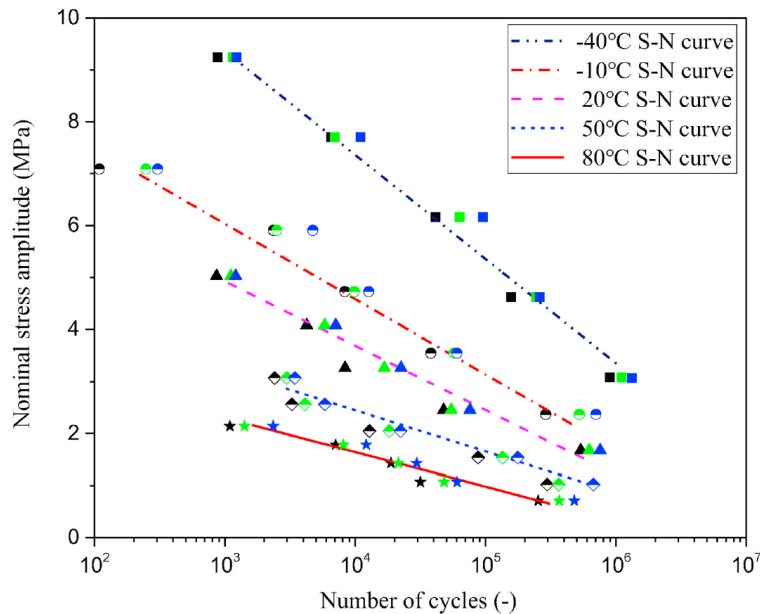


Figure 15: S-N curves at a range of test temperatures (reproduced from Figure 8 of Tan et al [28]).

These results are fitted to relationships in the format:

$$S = A \times \log(N_f) + B \quad (15)$$

S is the nominal stress amplitude in MPa, and N_f is the number of cycles to failure at this stress amplitude. The coefficients A and B are temperature-dependent. A polynomial fit to temperature is presented in Equation 3 of Tan et al [28], reproduced below. This formulae is used in the physical model to relate the calculated stress amplitude to a number of cycles to failure, given the calculated joint temperature, T .

$$S = (-1.41 + 1.31 \times 10^{-2} \times T - 4.13 \times 10^{-5} \times T^2) \times \log(N_f) + (10.11 - 0.11 \times T + 4.42 \times 10^{-4} \times T^2) \quad (16)$$

This experimental data is more representative of the joint filler than the previous reference used in LV Predict 1 [9], due to the use of the same adhesive type (PU) and more representative loading conditions (shear). As such, the physical model now focuses on fatigue crack initiation, rather than growth, as this is where the best experimental data is available.

For each stress cycle, the cycles to failure is calculated by rearranging Equation 16. This N_f value corresponds to the number of cycles to failure if the material were under continuous fatigue loading at this amplitude. The reciprocal of this, $1/N_f$, indicates the fraction of fatigue life consumed by this cycle. Fatigue damage for the joint is the sum of these increments of fatigue life use.

3.3.3 Plastic damage in polymer joint material

As for creep and fatigue, the plastic damage calculations in the physical model have been updated in reference to the new FE analyses discussed in 3.2.

From this FE, the relevant stress for plastic damage is the same as for the creep model: the maximum principal stress close to the region of stress concentration in the joint filler. As such, the same equation for stress is used (Equation 9 in Section 3.3.1.1). This stress is compared to the PU yield stress, taken to be 2 MPa, to determine if plastic strain is accumulated in the current stress cycle.

Plastic damage is calculated by comparing accumulated plastic strain during stress cycles to a maximum allowable plastic strain value, set at 0.13 (13%). The cycle estimation methods are used as described in two previous sections, utilising the improved temperature predictions described in Section 3.1. Plastic damage in other materials is not included in the updated damage model; plasticity is not a primary concern for older-type joints using bitumen filler material.

3.3.4 Chemical degradation of polymer insulation

Section 4.3.2 of LV Predict 1 Final Report [9] presented a risk analysis of potential chemical degradation mechanisms, concluding that several were sufficiently low risk to exclude from the modelling in this work. The exception was the reduction of mechanical properties due to time at high temperatures, which was recommended as an area for future iterations of the damage model to include. The risk analysis concluded that the properties of XLPE are likely to start to degrade after 20 years at 95°C.

As for the creep process discussed in Section 3.3.1, the relationship of chemical degradation rate to temperature is defined by an Arrhenius equation as shown in Equation 17.

$$Rate = A * \exp \frac{-Q}{RT} \quad (17)$$

The activation energy, Q (in J/mol), is specific to the material and physical process. Its value compared to RT , the product of the molar gas constant (8.31 J/K/mol) and the absolute temperature, dictates the exponential

increase in process rate with elevated temperature. As in the creep case, the constant A is assumed to be temperature-independent.

Equation 17 can apply to a known reference lifetime L_{ref} at a recorded experimental temperature, T_{ref} . After this time L_{ref} at a constant temperature T_{ref} , the damage is taken as 100%.

This gives:

$$\text{Rate} = \frac{100\%}{L_{ref}} = A * \exp \frac{-Q}{RT_{ref}} \quad (18)$$

For a more general creep dwell of duration t_p , at a constant temperature T_p , Equation 17 becomes:

$$\text{Rate} = \frac{d_p}{t_p} = A * \exp \frac{-Q}{RT_p} \quad (19)$$

By substituting the former expression into the latter, a general expression for damage accumulation during a single creep dwell is obtained:

$$d_p = t_p * \frac{1}{L_{ref} * \exp \left(\frac{Q}{RT_p} - \frac{Q}{RT_{ref}} \right)} \quad (20)$$

By summing the damage accumulated across the temperature time series, a corresponding degradation time series is produced, with 100% damage corresponding to failure via this mechanism.

Therefore, in the calculations described above, the rate of accumulation of damage is assumed to be time-independent (i.e. constant during a dwell at constant temperature). Figure 16 illustrates how the total damage would evolve over a temperature history applying the assumptions above.

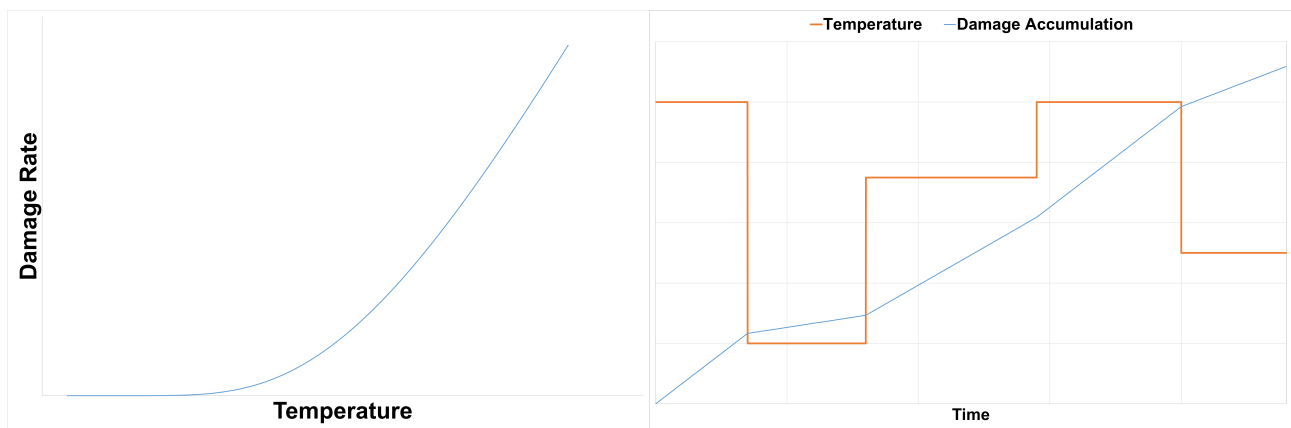


Figure 16: Illustration of a temperature-driven damage rate, and how the model uses it to produce a cumulative damage time-series from an input temperature time-series.

Experimental data from accelerated tests at high temperatures suggests an activation energy, Q , of 160 kJ/mol for the thermo-oxidative ageing of XLPE insulation [29]. This data is combined with the conclusion from the sensitivity

study during LV Predict 1 (assumed functional lifetime (L_{ref}) of 20 years held at 95 °C (T_{ref})), to provide a complete set of inputs for the damage calculation of Equation 20.

3.3.5 Chemical degradation of oil-soaked paper insulation

The insulating material of the older PILC cables is also expected to degrade chemically through life at the rated cable temperatures. Observations of failed cables [2–4] show examples drying of the paper and waxing of the oil compound, which increase the risk of arcing between conductors.

The damage of the paper insulation is defined by the same law as that of the polymer insulation, with appropriate changes to coefficients to reflect the different physical process and material properties.

The literature on paper insulation thermal behaviour is less comprehensive than for XLPE, due to its decreasing use for LV assets. However, accelerated thermal ageing tests and derivation of Arrhenius parameters has been attempted [30, 31]. The activation energy suggested in the literature [30], 111 kJ/mol, is assumed for this model. A functional lifetime (after which mechanical properties very rapidly degrade) is assumed to be 40 years at a constant temperature of 65 °C (the rated maximum conductor temperature for PILC cables, from Paragraph 3.5 of SD8B/4 [32])

Figure 17 presents illustrates the two chemical degradation models, by plotting the expected functional lifetime of each type of cable when held at a constant temperature. The plotted temperatures are normalised against the maximum conductor temperature listed in Paragraph 3.5 of SD8B/4 [32]: 90 °C for XLPE cables, and 65 °C for PILC cables. Note how, for both materials, the expected lifetime increases rapidly below around 80% of the maximum conductor temperature; thermal degradation is minimal except at very peak temperatures, as noted in Section 4.3.1 of LV Predict 1 Final Report [9]. Using the assumed Arrhenius parameters, XLPE sees greater acceleration of degradation at temperatures above nominal compared to the paper insulation - but thermal damage appears to affect the paper for a greater range of temperatures below nominal.

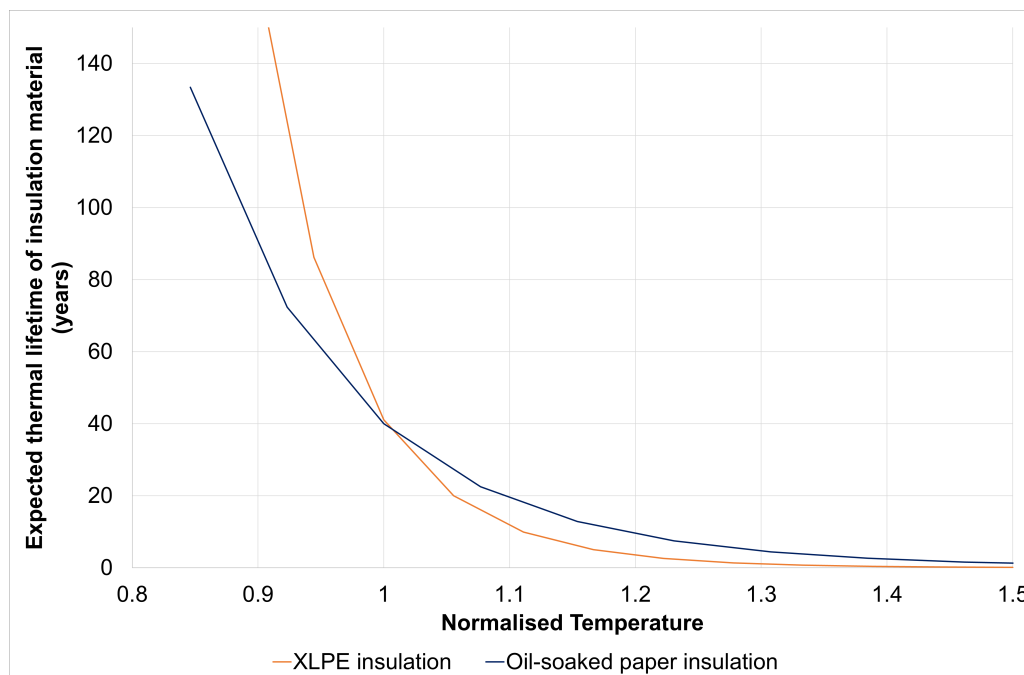


Figure 17: Expected lifetime of cable insulation under elevated temperatures - normalised against rated maximum conductor temperature for the relevant cable type.

3.3.6 Abrasion of paper insulation

Case studies of faulted PILC cables showed evidence of mechanical (wear) on some paper insulated conductors, likely due cyclic thermal expansion and contraction [2]. A particular risk area was identified by the investigators as the region in a joint where the conductors are splayed to allow connections. Resin fills the interstices in these areas, and so it is possible that cyclic differential thermal expansion can abrade the conductors' insulation.

However, other PILC case studies found no evidence of abraded paper insulation, even in older cables (insulation tapes from 1965 [3] and 1966 [4] showed no signs of abrasion, while the 1976 cable [2] did). Without access to the temperature histories of these cables, it is difficult to establish a link between thermal cycles and abrasive wear on the paper insulation. As noted in the SYNAPS 2 NIA project [2], the site of abrasion was where resin existed in the interstices of the conductors, where they are splayed for service connections. Therefore, the workmanship of an individual joint is likely a major factor in its vulnerability to this damage mechanism. Thus, even if the thermal histories of the cables [2–4] were known, there is insufficient data to distinguish temperature effects from joint-specific geometric effects.

The damage accrued by abrasive wear on the paper insulation would be expected to scale linearly with a number of thermal cycles - corresponding to a distance of sliding between adjacent conductors and/or filler resin:

$$\text{Wear (mm/cycle)} = C\Delta T \quad (21)$$

The coefficient C would be influenced by the conductor, insulation, and filler materials, alongside geometrical considerations like the distance between conductors and distribution of resin filler. For this work, there is insufficient literature on the abrasion of the paper insulation to define a robust damage calculation to input to the physical model.

Hence a necessary assumption is adopted; that damage component corresponding to abrasion of paper insulation is negligible.

3.3.7 Loss of pre-load of metal electrical connector bolts leading to runaway contact resistance

Joints in this model are assumed to use Insulation Piercing Connector (IPC) blocks using shear off bolts (see Figure 18). As discussed in Section 3.1.4, these are a source of additional heat input, and are likely to experience some of the highest temperatures and temperature ranges in the joint.

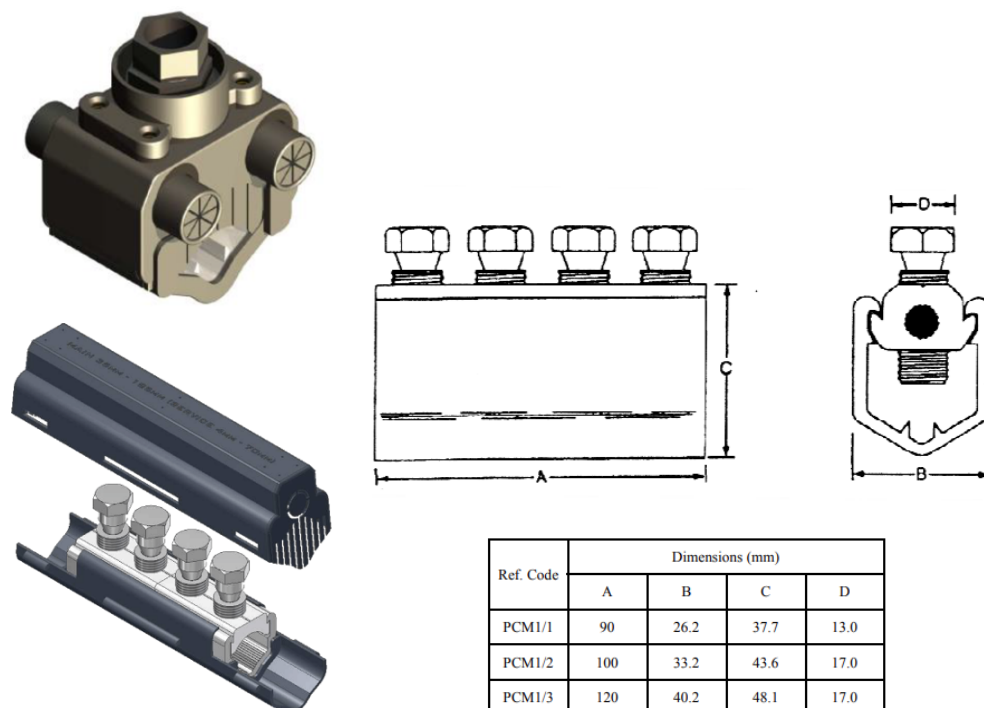


Figure 18: Examples of Insulation Piercing Connectors (IPC) assumed for all modelled joints [33–35].

The quality of the electrical connection relies on the load in the bolt, which is controlled by the shear-off of the bolt head during assembly at a rated torque. At high temperatures, relaxation of this pre-load is expected. This could lead to failure through an increase in contact resistance, causing higher temperatures. This can become a positive-feedback loop of increased temperatures, relaxing bolts, and increased resistance, and is hence named runaway contact resistance.

The compressive force on the conductor is calculated from the torque required to shear off the connecting bolt at its collar, taken to be 60% narrower than the head diameter (measured from the drawings in Figure 18). Assuming a shear stress for aluminium alloy 6061-T1 of 207 MPa [36], a shear off torque of 18.2 Nm and 40.6 Nm is required for the standard 13 mm and 17 mm IPC bolts used for joints. Assuming a typical bolt friction value, k , of 0.2, this corresponds to a compressive stress in the bolt of approximately 65 MPa (assumed to act over thread diameters 12 mm and 16 mm respectively). This is approximately 20% of the Ultimate Tensile Strength of the assumed aluminium alloy, A6061-T6.

The creep properties of the aluminium alloy can be used to predict the relaxation of this initial bolt stress, and the corresponding increase in contact resistance. A creep model from [37] links creep strain rate to stress and temperature:

$$\dot{\epsilon}^c = A[\sinh B\sigma]^n \exp \frac{-Q}{RT} \quad (22)$$

Where:

A =	1.05E15/hr	Creep strain model parameter [37]
B =	0.0404/MPa	Creep strain rate stress coefficient [37]
n =	1.74	Creep strain rate stress exponent [37]
Q =	170 kJ/mol	Activation energy for creep in the Al alloy [37, 38]
R =	0.00831 kJ/K/mol	Molar gas constant
T =	Joint temperature (taken from previous timestep), in Kelvin	

The bolt is loaded at a constant strain within the IPC, with the stress relaxing over time. Therefore, the accumulated creep strain is equal to the drop in elastic strain (i.e. an elastic follow-up factor of one). The relaxation of the bolt stress is then defined by this expression, based on the initial bolt stress (the shear off value) and the accumulated creep strain:

$$\sigma = \sigma_{t=0} - E \epsilon^c \quad (23)$$

Where:

$\sigma_{t=0}$	is the initial bolt stress
E	is the elastic modulus of Al alloy (70 GPa) [36]
ϵ^c	is the creep strain accumulated

The bolt pre-stress is small enough that complete relaxation of contact force on the conductor is unlikely during the timescales of interest, given the expected connector temperatures (potentially reaching twice the remote cable conductor temperature (rated 90 °C for modern cables)) . The more credible route to failure is the increase in contact resistance.

As noted in Section 3.1.4, the additional contact resistance can be directly related to the contact force from the connector bolt. Equation 7 and Table 2 of PD IEC/TR 60943 [12] include contributions from both constriction resistance and film resistance (the first and second terms respectively of this equation):

$$R_c = k_1 F^{-0.6} + k_2 \sigma_0 F^{-1} \quad (24)$$

Where:

k_1	= 1.3E-04	Constriction resistance constant for Al (Table 2 of PD IEC/TR 60943 [12])
k_2	= 1.35E08	Film resistance constant (Table 2 of PD IEC/TR 60943 [12])
σ_0	= 1E-10	Tunnel Resistivity of Al (Table 1 of PD IEC/TR 60943 [12])
F	Contact force from the connector bolt (N)	

Through the equations above, the feedback loop of temperature, bolt relaxation, and contact resistance can be used to calculate a more accurate temperature time series for the joint, as shown below. The higher temperatures

caused by considering contact resistance increases will be the input to all other damage mechanisms and will have an impact on rates of all failure modes.

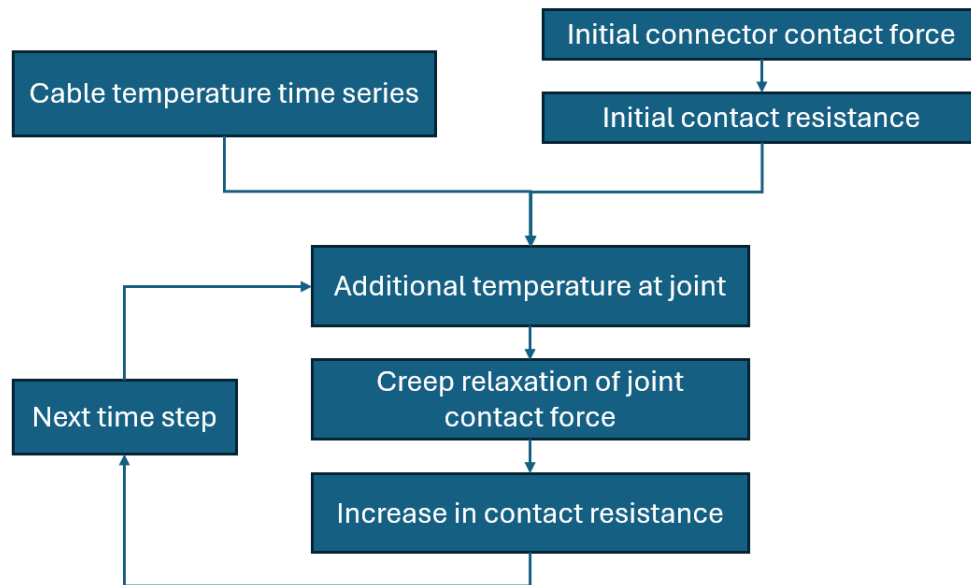


Figure 19: Illustration of the feedback loop between temperature, connector bolt load, and contact resistance.

Figure 20 shows the predicted increase in contact resistance for three test calculations, with demand assumed to be at a constant low/moderate/high level throughout lifetime. The coefficients in the equations above have been informed by literature and then tuned to ensure the understood behaviour is calculated; Figure 20 shows joints with low demands and temperatures see no change in contact resistance, due to limited creep in the aluminium bolts. The moderate demand test calculation shows a steady increase in contact resistance, with a factor of >3 increase from the initial value by 40 years of life. Runaway contact resistance induces very high temperatures in the joint model and therefore becomes extremely limiting of joint lifetime if operation is continuously at or above conductor rated temperatures.

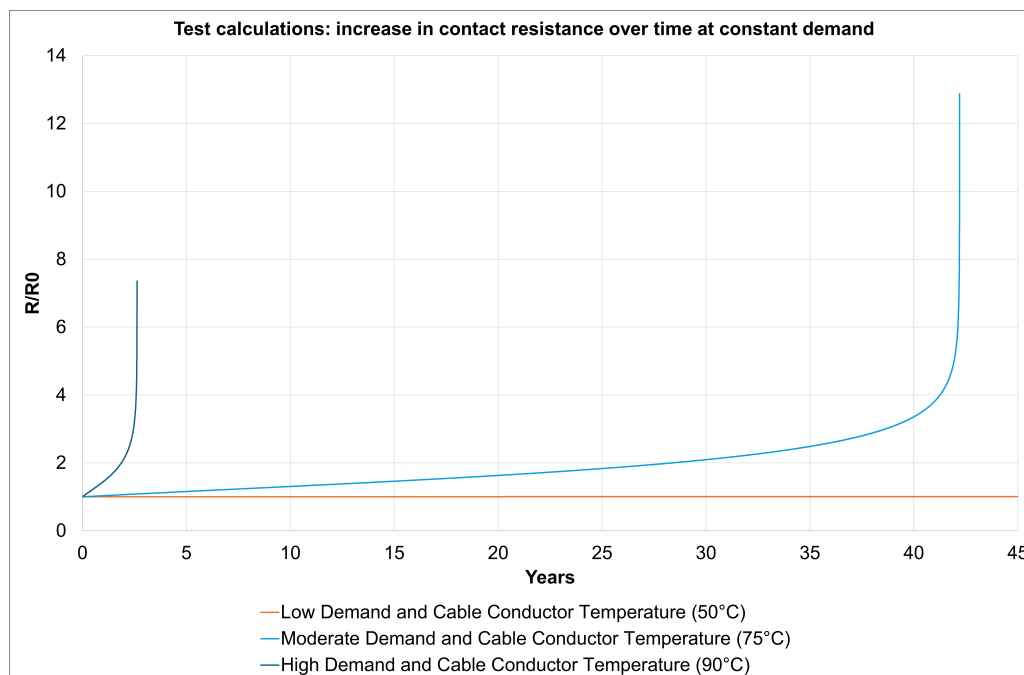


Figure 20: Evolution of contact resistance for three test calculations assuming constant low/mid/high levels of demand.

3.3.8 Fatigue of metal electrical connector bolts

Bolted connections are considered likely sites of fatigue failure for most equipment, particularly those operated at elevated temperature. It is possible that the thermal cycles applied by the lifetime loading profile in the joint causes cyclic stresses and the growth of fatigue cracks. This has been investigated for the IPCs described in Figure 18.

The cyclic thermal stresses will depend on the differential thermal expansion between the IPC material and the surrounding joint filler (either polyurethane or bitumen), when temperatures vary from the installation temperature (assumed 10 °C [32]). Both fillers have a greater coefficient of thermal expansion than aluminium, suggesting that the bolt is likely to see mostly compressive stresses during expansion, and its own expansion will not be constrained by non-expanding filler. In the bitumen case, high temperatures are also likely to soften the filler and thus the bolt will be more free to expand. Furthermore, many IPCs are a shrouded design, with a plastic casing between the bolt heads and the filler material. This protection from thermally-expanding filler, plus the clearance between the bolt (with head sheared-off) and the shroud, suggests that the IPC bolt will not see very large thermal stresses and is reasonably free to expand and contract with temperature.

Figure 21 shows an S-N curve for the assumed bolt material, suggesting an endurance limit of approximately 130 MPa stress range. This is double the predicted starting compressive bolt stress induced by the shear off during assembly (around 65 MPa). Although this source recommends lower values than the 130 MPa for design stresses [39], this assumes design factors of safety not directly applicable to this application. Figure 21 also assumes fully reversed loading conditions, so this compressive mean stress is not accounted for. It is not considered credible that thermal loading on the bolt would cause sufficient cyclic tensile stresses to exceed the endurance limit, based on the materials and constraints discussed above. In addition, to cause this tensile stress, the compressive assembly stress on the bolt would have to be completely unloaded, which would have a significant impact on the contact resistance. This would likely be detectable in the electrical signals of any monitored circuits; there are no reports

of daily signal changes indicating the connector bolt fully unloading and re-loading as demand cycles through daily peaks and troughs. Also, from the limited examinations of failed cables, no cracking or degradation of connector bolts have been reported, with some accounts noting the good condition of the metal connectors [2–4].

The materials data, geometrical constraints, and existing compressive stresses suggest that fatigue crack growth is not a significant failure mechanism for the buried LV cable joints. As such, the updated physical model does not calculate a damage metric for this. The assembled materials data and estimates of existing stresses could be used in future to derive a damage metric, should further cable inspections uncover evidence of fatigue crack growth in the metal connectors.

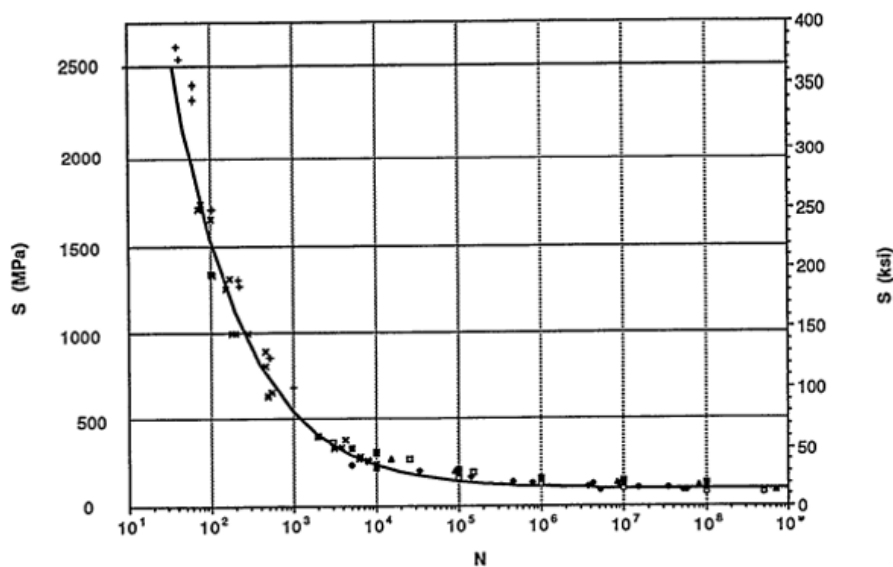


Fig. 1 Fit to fully reversed 6061-T6 fatigue data.

Figure 21: Fatigue data for the assumed bolt material [39].

3.3.9 Corrosion of lead cable or joint sheaths

Case studies of faulted PILC cable joints showed oxidation of the lead used as cable shielding and as the inner sheath of joints [3]. This lead seals the oil-impregnated cable components against external moisture. Some water was present within the joint sheaths, despite the bituminous filler of these older joints being understood to provide some 'healing' barrier against water ingress at cracks in the joint. Therefore, degradation of the lead components in older cables and joints may be linked to failure, either as a direct cause or as a symptom of other degradation or electrical faults.

The original choice of lead for sheathing for PILC cables was driven by lead's relative corrosion resistance, alongside its electrical properties suitable for the required common neutral/earth (CNE) connection. Lead in this application is understood to corrode reasonably slowly, benefiting from a passivating layer that slows the corrosion rate [40].

The oxidation of the lead sheath is a thermally driven process. Therefore, the greatest degradation of the lead will occur when it carries current, causing ohmic heating, i.e. on occasions of phase imbalance. It has been assumed for

this work that degradation is negligible when the sheath has no current flow. As discussed in Section 3.1.5, phase imbalance is a potential area of further methods development but currently not accounted for in the physical model. As such, degradation mechanisms linked to it are also not modelled. Literature on corrosion of lead suggests it is unlikely to be a direct cause of failure during PILC lifetimes, although some degradation is expected. This matches the observations of case studies, where corrosion of lead sheathing could not be directly linked to the electrical transients between conductors which would be the ultimate cause of failure; shorts between conductors were caused by degraded paper insulation, rather than external water ingress.

3.3.10 Fatigue failure of lead cable or joint sheaths

Beyond oxidation, it is likely that thermal loading contributes to development of cracks or leaks in lead sheath components. As noted above, the most extreme loading of the CNE lead will be during phase imbalances, where ohmic heating in the lead will cause thermal stresses. This loading is very likely to be cyclic, with periods of phase imbalance interspersed between normal balanced conditions where the lead carries no current.

As for the previous damage mechanism, fatigue failure in the CNE lead of PILC cables is highly linked to phase imbalance behaviour, which is as currently not in the scope of the damage model. Future work that models imbalanced loads should review these related failure mechanisms. For the current damage model, it has been assumed that damage to the lead sheathing material is not a direct cause of failure in cables and joints with balanced loads.

3.4 Grouping mechanisms for an updated damage model

At different locations in a joint, several of the finalised mechanisms above will be present and contributing to an overall likelihood of failure. The damage formula from each mechanism has been defined such that summing the damages provides a total damage metric, which corresponds to failure when it reaches 1.

A review of the limiting damage locations was informed by the visit to the ENW Training Academy and by the observations (from the SYNAPS 2 NIA projects) of the sectioned cables [2–4]. This confirmed that the chosen location from LV Predict 1 was reasonable; the interface between the cable and the joint filler material is vulnerable to the degradation mechanisms of the insulation, the filler, and any defects that could grow between them. For this work, this location is defined as 'damage check location 1'. A second location is in the vicinity of the electrical connection between conductors. This is the hottest location in the joint as it is the source of the additional ohmic heating introduced for LV Predict 2, and will also see contributions from the different types of insulation on the conductors being connected. This is defined as 'damage check location 2' for the joint.

Table 5 illustrates how the joint's attributes are constructed in the physical model method, and the damage contributions at each damage check location. The structure of the model considers the properties of the cables entering the joint. If XLPE cables are present (even just one), the model assumes that the more modern resin filler is cast into the joint. If only PILC cables exist, a bitumen filler is assumed. As noted in Section 3.3.7, we assume they are all aluminium IPCs.

At check location 1, the mechanisms defined in LV Predict 1 are modelled, using the updated LV Predict 2 temperature modelling. The new mechanisms of chemical degradation are added to the LV predict 1 mechanisms in this location. Conservatively, the model assesses the damage from all these mechanisms at the maximum

joint temperature (the cable temperature plus the effects of the joint insulation and the bolt contact resistance), as calculated in Sections 3.1.4 and 3.3.7.

At check location 2, the insulation damage mechanisms (creep and chemical degradation) are combined, considering both conductors at the connection separately. This location also uses the maximum joint temperature (the cable temperature plus the effects of the joint insulation and the bolt contact resistance), with the new consideration of increasing contact resistance due to connector bolt creep.

For the modern polymer-filled joints damage check location 1 is limiting, based on the number of mechanisms and the use of the joint maximum temperature. For bitumen-filled joints, the mechanisms related to polymer filler (plastic damage and fatigue crack growth) are not applicable, and the damage sums for both locations become identical.

Therefore, for the damage mechanisms quantified in this work, each joint filler type has a consistent single damage check location. However, this updated model structure allows for checking multiple locations easily if future modelling work incorporates other mechanisms such as those discussed but discounted above. The review work of modelling assumptions and the operational experience of real joint failure locations highlights that this is a valuable structural change to the physical modelling methods.

The addition of the chemical degradation to the damage total would lead to more pessimistic results compared to an analysis using the LV Predict 1 PoC model. Likewise, the incorporation of local heating at joint connections, and the feedback loop of loss of bolt pre-load, means a cable modelled with the LV Predict 2 model would see higher peak temperatures than if it were modelled with the PoC model. However, considering transient temperatures has been shown to significantly reduce these peaks from their steady-state values, generally giving a more moderate temperature history. These effects will compete to different extents depending on cable/joint type, and on volatility of their demand history (high temperature effects vs. cyclic thermal loads).

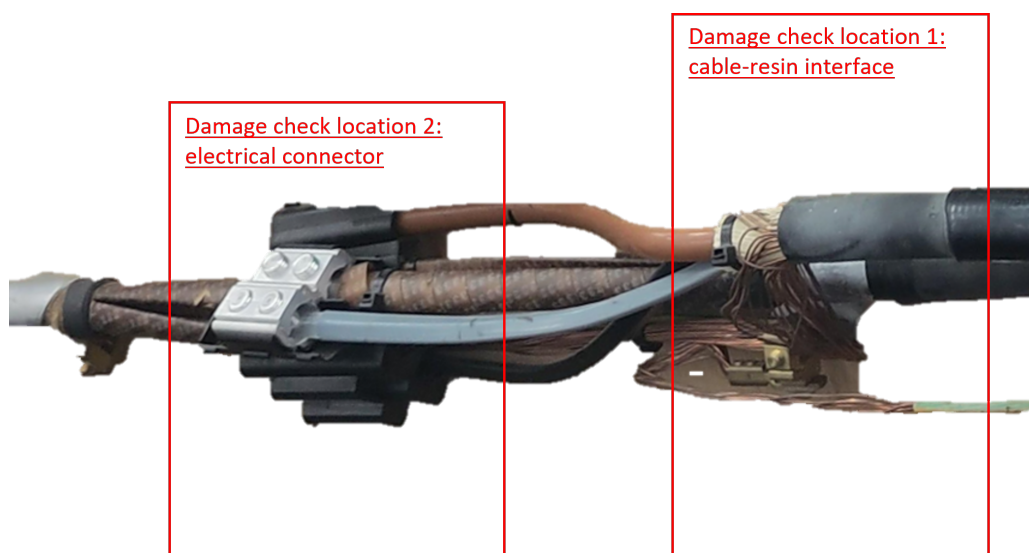


Figure 22: Finalised two damage check locations, where contributing damage mechanisms are tracked and summed.

Joint Object		
Components	Materials	
Cable 1	PILC or XLPE	
Cable 2	PILC or XLPE	
Joint Filler Material	If both cables PILC: Bitumen If some XLPE: polymer	
Electrical Connections	Assume all are Aluminium Shear-off bolt IPCs	
Damage check location	Temperature history source	Damage contributions
Cable/joint filler interface	Joint peak temperature (including effect of contact resistance increases)	If polymer filler: polymer filler and insulation creep; polymer filler plastic damage; fatigue crack growth; polymer insulation chemical degradation; paper insulation chemical degradation If bitumen filler: paper insulation chemical degradation only
Conductor connections	Joint peak temperature (including effect of contact resistance increases)	Polymer insulation creep; polymer insulation chemical degradation; paper insulation chemical degradation

Table 5: Summary of how joints are constructed and checked for damage in the physical model.

3.5 Physical modelling assumptions, judgements, and limitations log

This section lists the major assumptions and judgements made in the updated physical model, highlighting significant known limitations. Several suggest areas of future research and model development.

1. **Imperfect experimental data:** experimental data has been used to derive material properties and empirical relationships [2–4, 23, 25, 28, 29, 31, 36, 37]. There is inherent uncertainty in these estimates. As discussed in the previous sections, some properties have been scaled or tuned to be suitable for this work; this adds additional uncertainty. Further data, targeted specifically at buried LV cable and joint materials, would improve confidence in the properties assumed.
2. **Only conductor electrical resistivity is temperature-dependent:** the temperature-dependence of other material properties is judged to be insignificant.
3. **Joint filler type:** the joint filler is assumed based on the joint's cable type, as shown in Table 5. It is expected this will be correct for the majority of joints, but there will be a minority of joints consisting of only PILC cables that use modern polymer filler, for example in the case of a repair. This assignment governs which damage mechanisms are active in the model.
4. **Connector type:** all connections in cable joints are assumed to be made using aluminium shear-off Insulation Piercing Conductors (IPCs).
5. **No phase imbalance:** no phase imbalance is considered in the scope of this updated model, although investigation suggests it could be an important factor for degradation.
6. **Backfill with different soil grades is ignored:** when joints are installed or fixed, their excavation is often backfilled with a standardised grade of sand/soil. This means that regional data on soil types may not accurately reflect the soil in direct contact with cable joints.
7. **The model links current to temperature via simplified expressions:** 2D and 3D FE models were used to inform the formulae and tune certain parameters, but are not used directly in the model.
8. **A simple factor accounts for the difference in insulating effect between cable insulation and the joint filler and casing:** the factor balances the greater thickness of the joint filler and its different thermal conductivity compared to the cable insulating material. This factor would be validated using experimental data on the joint filler material, which is currently not available.
9. **An empirical factor accounts for the area of soil impacted by the cable temperature:** f_{soil} , used in the transient temperature calculations, is based on the FE studies in this work and is not generally applicable outside this context.
10. **Joint peak temperature is used for all damage mechanisms:** conservatively, all damage mechanisms use the maximum joint temperature (i.e. the cable temperature plus the effects of the joint insulation and the bolt contact resistance). In reality, this maximum temperature decays down to the remote cable temperature away from the bolt contacts; the far edges of the joint may be significantly cooler than their assumed temperature in the model.
11. **Total damage is calculated as the sum of the separate mechanisms:** damage is accumulated for each mechanism via summing fractions of lifetime used up by each timestep. A simple sum of all relevant damage

mechanisms (depending on cable and joint type) is taken as the 'total damage' measure for the joint. In reality, the interaction between mechanisms, and their corresponding triggering of failure (see below), is more complex. Modelling this interaction would require more experimental data including examination of a large sample of pre- and post-failure joints, and some corresponding monitoring data.

12. **Total damage is not an exact predictor of failure time or failure mode:** Damage becoming equal to 1 does not necessarily correspond to failure of the joint at that time — both due to the uncertainty in summing damage but also the steps between damage and electrical failure. For example, damage may accumulate enough that a joint is vulnerable to water ingress, but failure may not occur until a period of wet weather occurs.
13. **Imbalance-related damage mechanisms are excluded:** the temperature model assumes balanced phases, therefore imbalance-related degradation mechanisms are ignored (e.g. thermal degradation of CNE lead sheaths).
14. **Electrical damage via partial discharge is ignored:** partial discharges are not a credible electrical-based degradation mechanism — the only electrical-based degradation is through water ingress causing “dry band arcing”, accounted for in the other degradation mechanisms.
15. **Other chemical degradation mechanisms are ignored:** aside from the degradation of polymers due to long dwells at high temperatures, no chemical degradation is assumed (e.g. leeching).
16. **Damage rates are time-independent within each dwell:** the rate of damage accumulation during a single temperature dwell is assumed to be constant, as illustrated in Figure 6. In reality degradation rates change over time even without temperature changes.
17. **Starting connector bolt stress:** The shear-off IPCs for all connections are assumed to give a starting compressive bolt stress of 65 MPa.
18. **Fatigue failure of connector bolts assumed to be precluded by their starting compressive stress:** fatigue failure would occur via cracks opening during the tensile portion of any cyclic loading. The pre-stress of the bolts is judged to be great enough that no significant tensile load applies.
19. **Joints are assumed to be defect-free at start of life:** fatigue in the joint filler material is based on fatigue damage initiation. As there are many possible locations for an initial defect, no single location can be reliably assumed for all modelled joints.

4 Network simulation: demand to damage

This chapter outlines the entire process of simulating network conditions, from sampling demands to calculating the damage in joints that these demands would result in. Several developed models within this process are then discussed in more detail in their respective chapters, including the conversion of demands to currents through load flow emulators (Section 5.2). This chapter also includes explanation of how the physics-based models for joint temperature and damage (Chapter 3) are utilised to generate training data.

Demands can vary greatly between households, not just in overall energy consumption but also in the profile of energy use throughout the day, week and over the course of the year. Aggregation to different numbers of customers, and with a different set of households, will produce different demands on the network. To produce a dataset that represents the underlying distribution of demands for a given group of customers of a certain size and on a certain network topology, sampling of demand time series must be repeated many times. This process of picking samples randomly, many times, is called Monte-Carlo (MC) sampling. This is a common approach when the underlying distribution is not known, or when it cannot be represented with a fixed shape or class of distribution (such as a normal distribution). In this case, sampling of entire time series also allowed us to retain the relationships and correlations between time points that is seen in real data, an important feature that will affect temperature cycles and therefore damage in the cables.

The random variation in demands between individual households is not the only variability we need to capture over the whole simulated dataset; several other variables are assumed to be known, but can vary between cables and joints. Network topology, the mix of cable types and geographic location (including elevation, soil types and rural vs urban areas) all vary across networks as well as variation in LCTs installed under different scenarios. This sampling across multiple dimensions of varying parameters greatly increases the number of samples needed for proper coverage of all possible outcomes. The factors that were varied in the creation of the whole sampled (and then simulated) dataset, and the number of samples drawn for each run, is summarised in Table 6. This shows how rapidly the number of samples needed increases as new factors are added, quickly increasing the computation time needed for this process. While this table shows the size of dataset created during this project, these numbers are not fixed. In particular we would recommend increasing the number of networks sampled to get greater coverage of varying network properties across ENW's entire license area.

The previous LV Predict project [9] simulated a set of cables with a range of properties and demands, treating every cable as if it had a single joint. While this methodology was sufficient to demonstrate the proof-of-concept and create a set of reasonably realistic cables with simulated demands, it did not fully reflect the variety of topologies and realistic loads seen across ENW's LV network. In this work, we have extended the physical modelling and simulation of cable demands and currents (which were then used as the inputs to calculations of temperature and damage) in several ways:

- ▶ Modelling of network topologies, taken from Geographic Information System (GIS) data for ENW's network. Connected network downstream of a secondary transformer was modelled as one network, and networks connected to different transformers were considered independently. Cable properties were extracted from GIS, and actual recorded numbers of customers connected at each service point were used to form a set of real networks to simulate.
- ▶ Extended sampling: Sampling of a range of LCT penetrations for each service point with households connected, and inclusion of heat pump demands as well as Electric Vehicles (EVs).

Variable	How values are chosen	Number of samples
Secondary transformer	Random: from set of ENW transformers	111
Average # cables modelled	All cables around valid joints (>1 physical cable with customers connected)	17.9
Number of customers at each service point	Real values from aggregated MPAN counts	1
Number of EVs and heat pumps	Penetrations chosen randomly and scaled to number of customers	10
Specific set of demand series sampled	Randomly from set of possible series	3
Naïve vs smart EV charging	separate sets of time series	2
Total number of time series		119214
Time points	Half-hours in available dataset	51360
Total number of time points modelled		6.12 billion

Table 6: Number of samples generated across each parameter variation.

- Load flow modelling: sampled demands were supplied to a load flow model along with the network representation of all feeders connected to the given transformer, returning realistic currents at selected locations on the network.
- Emulators: to enable a large volume of simulations to be undertaken, emulator models (Machine Learning (ML) models) were built. Instead of running load flows for every single time point simulated, high volumes of predictions can be made from trained emulators for each cable.
- Improved, more realistic physical modelling of temperature: Additional effects such as contact resistance are included, and dynamic temperatures (instead of steady-state values) are calculated.
- Improved, more realistic physical modelling of damage. See Chapter 3 for full details on the temperature and damage models.

4.1 Sampling and simulation process

Figure 23 shows all the steps involved in the complete sampling and simulation process. The full sampling and simulation process is described step by step below, following the form of this flowchart. Steps 1, 4 (customer numbers only), 5 and 6 only need to be calculated once per network, so are conducted separately before the main sampling process is run and those pre-processed inputs are read in directly for efficiency. The input network properties table contains all physical information about the network: it has a row per cable and specifies the connecting busbars at either end, cable type, conductor area, length, and location.

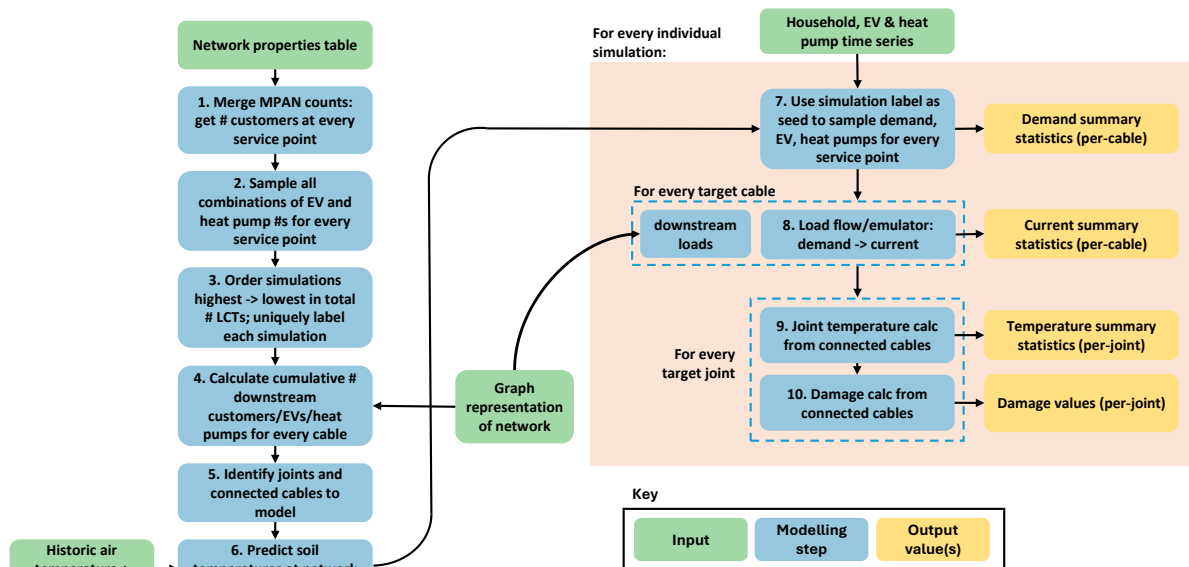


Figure 23: Schematic showing all steps in the sampling and physical modelling process.

4.1.1 Step 1: Collating customer numbers

ENW provided a dataset of the total number of customers (the number of unique Meter Point Administration Numbers (MPANs)) connected by service point, which was merged with the set of service points on the network being modelled.

4.1.2 Step 2: Sampling smart meter, EV and heat pump numbers

The true number of smart meters, EVs and heat pumps on each network is not accurately known, and the penetrations of these technologies are expected to change in future. Because of this, the sampling process was designed to produce samples containing a wide range of smart meter, EV and heat pump penetrations across all possible levels, including possible future levels. These samples were ultimately used to train the final statistical models, so it is important that they span the range of all possible conditions that we could reasonably want to make predictions for, to avoid extrapolation in the predictions. Penetrations were drawn from a uniform sample, with smart meter penetrations between 0 and 1 per household, EV penetrations between 0 and 2 per household (to account for households with multiple vehicle charge points), and heat pump penetrations between 0 and 1 per household. The process for assigning a number of smart meters (or EVs or heat pumps) to a cable with a given number of customers n is as follows:

- ▶ Sample a random number x_r , $0 \leq x_r \leq 1$
- ▶ Calculate the minimum and maximum number of smart meters allowed: $min_{sm} = 0$; $max_{sm} = n \times max_{pen}$ where $max_{pen} = 1$ for smart meters (it would be 2 for EVs or 1 for heat pumps).
- ▶ Convert x_r to a number of smart meters allocated: $n_r = x_r \times (max_{sm} - min_{sm} + 1) + min_{sm}$, rounded down to the nearest integer.

This formula ensures that, across a large sample, the distributions of smart meter, EV and heat pump numbers

are approximately uniform up to the maximum specified penetration levels. For example, a cable with one customer could have 0 or 1 heat pumps and each of those two options would be selected 50% of the time. By sampling random values x_r for heat pumps and EVs independently, the overall sampling set will include possible combinations of EVs and heat pumps across the whole range of both technologies. The set of random numbers x_r used to generate EV and heat pump numbers for all service points under all sample combinations is generated in one go, with a consistent random seed, the transformer ID for that network, for replicability.

4.1.3 Step 3: Sample ordering and labelling

Later in the process (in step 8), a small subset of the sampled demands are run through load flows run to create the emulator training dataset for that network, and this is only done the first time a simulation for that network is run. The range of demands included in this is based on the range of demands seen in the full set of sampled demands for that one simulation. Therefore, it is important the first simulation for a given network includes the highest expected demands that will be seen across all the simulations. For this reason, the sample with the highest number of total LCTs (EVs and heat pumps) must be the first sample to be run. Once EV and heat pump numbers have been assigned to the network, samples are ordered by decreasing total number of LCTs before a unique ID is assigned. This ensures simulations are performed iteratively, starting with the lowest unique ID (with the highest number of LCTs). The number of customers is not considered as this is constant for a given network across all simulations. To ensure the creation of emulator training data (load flow runs) and emulator model training is run only once for each network, all simulations for the same network are run sequentially on one core - and the code is parallelised over different networks to increase the speed of the overall process and make best use of the computational resource available. The unique ID is used as a seed for all subsequent sampling (picking the exact demand time series selected for each simulation) so the process is replicable but the demands selected are still random.

4.1.4 Step 4: Cumulative customer and technology numbers

The graph representation of the network, a "networkx" object in python, is used to compute the total number of customers, EVs and heat pumps downstream at every cable in the network. For each cable, this is achieved by iterating over every node with a connected load, and checking whether any valid paths between that load and the transformer pass through the cable of interest. If so, the number of customers (or EVs or heat pumps, depending what is being accumulated) at that node are added to the total number of downstream customers for that cable. The process is repeated for all cables in the graph to get a complete set of cumulative numbers across that network. The whole calculation is repeated separately for the number of customers (i.e. base household demands), EVs and heat pumps as the number of each is needed to sample the different demand time series for each, and as an input feature to the statistical models. While all the networks analysed in the sampling process were purely radial, this accumulation method is also valid for meshed networks. In this case, care must be taken to ensure no double counting of customers occurs if there is more than one path between the load and transformer: the *set* of cables making up *all* possible paths should be checked (rather than counting the same cable twice in two different paths).

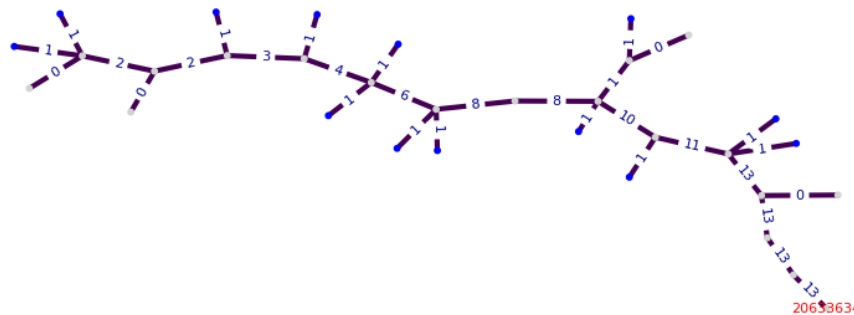


Figure 24: Example network graph with accumulated customer numbers on . Blue nodes are service connections with at least one customer connected. Red label denotes the secondary transformer.

An example of calculated customer counts for a small network is shown in Figure 24. This also shows some network branches or single cables terminating without a load attached, and this is common across all networks analysed. This is because the only loads modelled are those with at least one MPAN at the service point. Other end nodes without loads connected in this network model are likely unmetered connections to street furniture such as street lights and traffic lights. The demands from these assets are likely very small relative to domestic customers so unlikely to contribute meaningfully to the highest loads seen on the network. For this reason, and a lack of data on both locations and high-resolution demands making their contribution very uncertain, they have been excluded from the demand modelling.

4.1.5 Step 5: Identifying joints and cables to model

While it would be possible to fully model and record every single cable and joint on the network, there are some joints where modelling would not add value, for example where there are no customers connected to any of the cables feeding in and out of the joint. The joints to model are selected based on the following criteria:

- ▶ There must be at least two individual non-service cables connected to the joint. A service breaches joint (as opposed to a 'full' LV mains joint) occurs where a service cable connects to an LV cable, and this is assumed to be connected to one phase rather than a full join between two different cables. The contact resistance and therefore heating in a service breaches joint is likely significantly lower than a full joint. Additionally, service cables would only be replaced on demand and never proactively, so there is little benefit to modelling this type of joint for the purposes of this project.
- ▶ There must be a non-zero number of customers connected to at least one cable at the joint. While joints with no current flowing through them may still experience degradation over time, there will be no additional degradation driven by the thermal effect of currents and therefore this would be expected to always be lower than for joints with a current.
- ▶ The cable type and diameter of all connected cables must be specified. In this demonstration, we have limited modelled joints to those where the cable types and diameters are given in the GIS data. While this reduces the number of joints modelled, it ensures all joints that are modelled are physically accurate and avoids making assumptions about network properties in the modelling and simulation process. Since the main

aim of the physical modelling is to create a set of representative joints, it is not crucial to model all joints on every network. To increase the number of joints modelled per network in the physical simulations, it would be possible to use an estimate of the most likely values for missing cable types and diameters instead.

Once the set of joints on the network to model has been identified, the set of all cables connected to those joints is also identified to ensure currents are modelled for all connected cables around all joints of interest.

4.1.6 Step 6: Soil temperatures

A statistical model for soil temperature at the locations of joints was developed, and is used at this stage in the sampling process to generate soil temperatures for each joint location that are then supplied to the temperature modelling in Step 9. For the full detail of the development of these models, see Section 3.1.6.

For the original LV Predict project, predictions were saved out for random locations assigned to the set of simulated cables. However, in LV Predict 2, many more simulations are run with a much larger number of different cables and locations. Since the soil temperature model was developed in R, soil temperature time series for every modelled joint location per network were predicted and saved out separately prior to the main simulation process: the file of soil temperature time series for that network could then be read in at the appropriate point in the sampling process. If ENW were to develop this methodology into a useable tool it would be beneficial to implement all modelling in the same language to ensure seamless integration of all elements of the modelling.

4.1.7 Step 7: Demand sampling

Once all the preceding steps have been completed, the parameters for every simulation for the network have been set up. Each individual simulation (combination of LCTs and unique sample of demand time series) is conducted iteratively for one network, with other networks being run in parallel simultaneously. The number of parallel processes (and therefore the number of networks that could be simulated at once) was limited by the available Random Access Memory (RAM), as each process needs to hold a fairly large amount of data in memory. The full time series for all cables and joints being simulated, as well as the graph and network (IPSA) representations of the network, are required for every simulation. A random number generator is initialised with a unique seed for each sample for replicability, and then demand sampling follows a set of steps which are repeated for every service point on the network:

1. Read in the column names of the available demand time series. For each service point on the network, use the random number generator to sample n of these, corresponding to the number of customers connected at that service point.
2. Read in only the columns of demand data selected in the previous step (plus the datetime column).
3. Repeat the previous two steps with the EV and heat pump demand time series to get those demands for the service point as well.
4. Select the smart meter subset of the total selected demand — m smart meter columns are selected (without replacement) from the household demand set for each service point. The number of EVs and heat pumps in the smart meter subset is calculated from the number of smart meters multiplied by the overall EV or heat pump penetration for that service point, plus a random perturbation and rounded to the nearest integer. Similarly then the smart meter subset of EV and heat pump demands are chosen from the EV and heat pump

demand already selected for that service point, and summed with the household smart meter demands per time point to get the total smart meter demand for that service point.

5. The total demand (household + EVs + heat pumps) is now summed per time point, with the resulting series named with the service point ID.
6. Service point demand and service point smart meter demand are returned.

The set of service point demands is concatenated to produce a table with a row for every time point and a column for every service point. Two tables are produced in this way — one for the total service point demand and one for the smart meter demand at each service point.

4.1.7.1 Household demands

Half-hourly demand time series are sampled for a group of n customers by choosing a set of n demand time series at random, with replacement, and summing them. Demand data from the Customer-Led Network Revolution (CLNR) project was used [41], and there were several key considerations for selecting this dataset. The suitability of the input demand time series depends on:

- ▶ The length of the time series. At minimum, one full year of data is needed to capture seasonal variability in demand. However, multiple years of data is needed (and preferred) to also capture variability between years for the same households. The CLNR data is suitable in this regard as it includes three years of continuous half-hourly demands.
- ▶ Number of time series: the Monte-Carlo (MC) sampling method relies on having a sufficient number of individual samples to select demand time series from. The CLNR dataset includes smart meter data from around 9000 domestic customers, representing a range of households and including the possible variety is household demand behaviours seen in reality.
- ▶ Resolution of the data: Half-hourly resolution data is used throughout the process of simulating demands and currents. This time resolution of data does not include information on effects seen at much shorter time scales, such as transient effects in the network; however, it was determined that cable heating was the dominant driver of damage accumulation and this is the modelled phenomenon in this project. Due to the thermal inertia of the cables, half-hourly resolution is sufficient to capture temperature variations. Additionally, higher-resolution demand data is much less commonly available, whereas half-hourly data allows the possibility of substituting or augmenting the possible set of demand time series with other smart meter data in the future.
- ▶ Completeness of the data: several smart meter time series included long periods or significant portions of missing or faulty data. Time series with over 25% of timestamps missing were excluded from the sampling process, resulting in a set of 4719 possible smart meter time series to sample from.
- ▶ Age of the dataset. Electricity consumption changes gradually over time, due to changing consumer habits, appliances used and energy efficiency. The CLNR dataset covers 2011–2013; average domestic electricity consumption was 18.5% lower in 2023 compared to this period, according to the Government's reporting [42]. To best represent current energy usage, recent smart meter data would be most appropriate - but this would not reflect historic demands as closely and is difficult to access in a useful format.
- ▶ Pure household demand data, without EV and heat pump demands also being measured in the same quantity. Household demands, EVs and heat pumps were modelled separately in this project to allow simulation of

different levels of future technology adoption. A smart meter dataset where these technologies were present within the data but not explicitly specified would require much more additional processing. Even if EVs and heat pumps were specified, not having separate time series for these demand components would require careful reformatting of the whole sampling methodology and may limit the combinations of exactly which households could be used under different scenarios.

The inclusion of additional (and more recent) smart meter demand datasets with a larger number of households was investigated. Unfortunately, this was not feasible within the time and budget constraints of the project, but should be considered in greater depth if this tool is developed further in future. The CLNR dataset was selected as the most suitable and readily available source of smart meter data, but inclusion of smart meter data from other sources as well and particularly more recent data would be of real benefit.

All real datasets will have imperfect data; the most obvious example is missing values. In the CLNR dataset, 17.8% of half-hourly timestamps are missing across all retained time series: the distribution of proportion of missing values is shown in Figure 25. The physical models for temperature require complete input time series, so these missing values must be filled in an appropriate way. Within the sampling process, when a set of smart meters has been randomly selected, missing values are filled with the mean for the group across all complete time points for the same time of day and month of the year. This ensures the group aggregated demands are not distorted by the number of values available at each time point, while also maintaining the mean power seen across the available time stamps and preserving observed seasonal and daily profiles as much as possible.

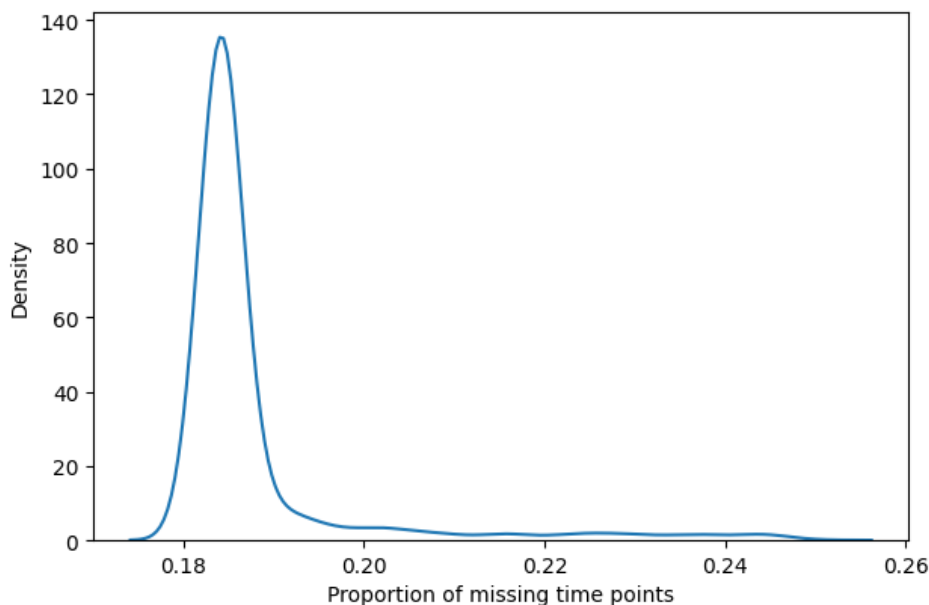


Figure 25: Distribution of proportion of missing values across all smart meters in the CLNR dataset.

The average daily demand profiles by day of week are shown in Figure 26. The morning increase in demand is shifted slightly later with higher demand throughout the daytime at weekends. However, the evening peak is very similar across all weekdays, which might be expected from purely household demand (since the majority of the reduction in demand on weekends comes from closed commercial buildings, not households).

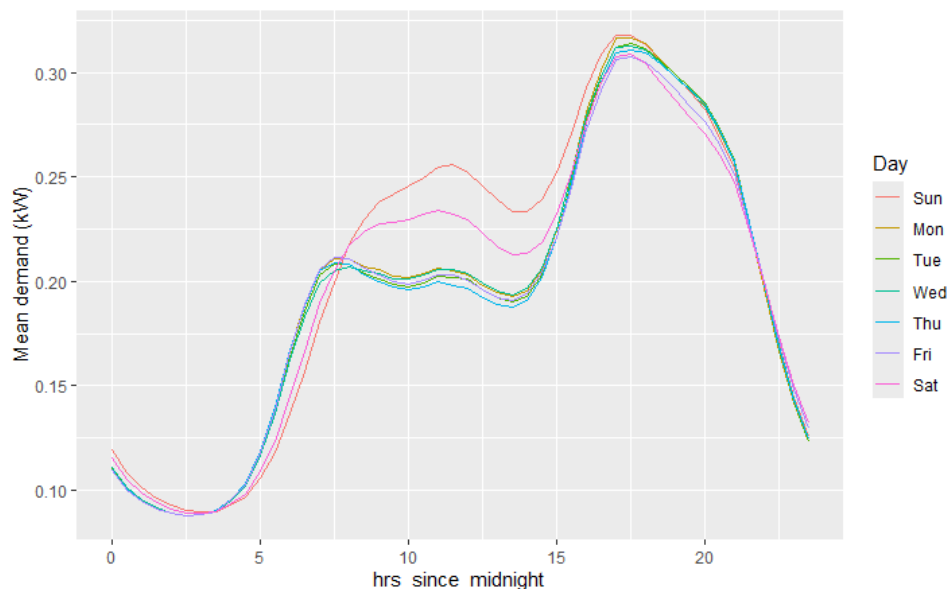


Figure 26: Average daily demand profiles across all smart meters in the CLNR dataset.

4.1.7.2 LCT demands

Customer demand time series from smart meters are assumed to represent the real range of demand behaviour seen across customers, but only in their domestic appliance consumption. While some of these time series include (resistive) electric heating demands, it is assumed that none include demand from heat pumps for heating, or electric vehicles for transport. Given the age of the CLNR dataset, this is a reasonable assumption. Demand time series for these two LCTs were modelled separately, allowing control over the penetrations of both EVs and heat pumps in each sample, and producing a set of samples with varying levels of both these technologies. Creating a dataset of cable conditions that spans a wide range of LCT penetrations is important to simulate and understand the effect of future demands, when much higher uptake of both EVs and heat pumps is expected. The final model will then be able to accurately predict cable degradation under future scenarios as its training set includes these conditions: this is a crucial capability to enable future network planning.

Processing of EV time series

EV time series were taken from the Electric Nation project [43], which recorded the plug in times and demands of electric vehicles over the course of a year (2017). The same data considerations were applied as for domestic demand data, although the available data for EV charging demands is smaller in both number of unique EVs and length of time available as available data from trials and studies on EV charging is more limited. The Electric Nation dataset includes some smart charging with no one unified tariff across the complete set of customers. Therefore, we processed this data to create two time series (an example shown in Figure 27):

1. Naïve charging. In this dataset, a charging session is defined as the timespan a given vehicle is connected to the charger for. The vehicles charge as soon as they are physically plugged in, and continue to charge at the rated power of the charger until the total volume of energy for that charging session is reached, beyond which point the demand is 0 until the start of the next charging session (the next time the EV is plugged in). The same total demand is seen as in the original recorded data, but charging occurs as soon as the car is plugged in, i.e. without any smart charging or flexible behaviour. This is expected to represent the worst case

scenario, where EV demand increases peak demand on LV cables, as often EVs are plugged in at peak hours when customers return home from work.

2. Smart charging, specifically a dual-price tariff model. For this scenario, all EV owners were assumed to only charge their EVs during off-peak hours. Off-peak hours were defined as midnight to 6 am, as these are the most common off-peak times seen across real tariffs [44]. EVs were assumed to begin charging at the start of the off-peak period, with total off-peak demand for the day matching the total demand for the whole day in the naïve charging time series. If the total demand exceeded what could be consumed at rated power over the entire off-peak period, the additional demand required was added to the next off-peak period for the following day. This method preserves the total demand of each EV, purely shifting the time of charging within the day but retaining daily and seasonal charging patterns.

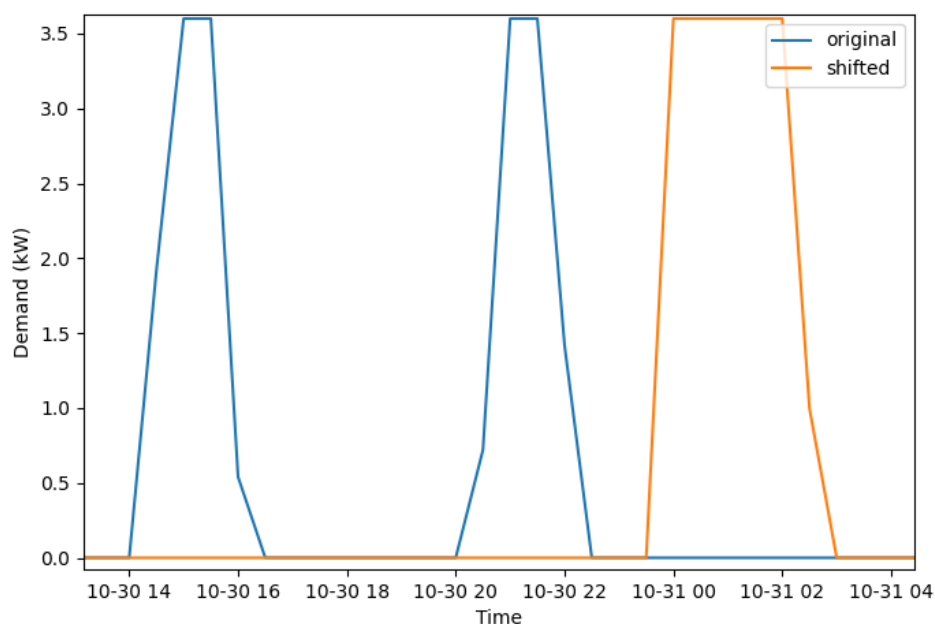


Figure 27: Example day of naïve and smart charging. For the smart charging scenario, all demand is shifted to the off-peak period. X-axis labels have the format 'month-day hour'.

The naïve and smart charging time series represent the two extremes of possible behaviours: either no smart charging at all, or only smart charging. Real charging behaviour is expected to lie somewhere in between these two extremes, but exactly where is very uncertain. Thus, modelling the two extremes, which each correspond to well specified scenarios, puts bounds on the possible effect of EVs. As shown in Figure 28, the average collective profile of EVs with naïve charging has a peak that coincides with the evening peak in domestic demand. The peak in the smart charging collective profiles is much larger, as all EVs that charge on a given day begin charging at the same time (the start of the off-peak period). However, this charging is shifted to overnight when other demands are lower.

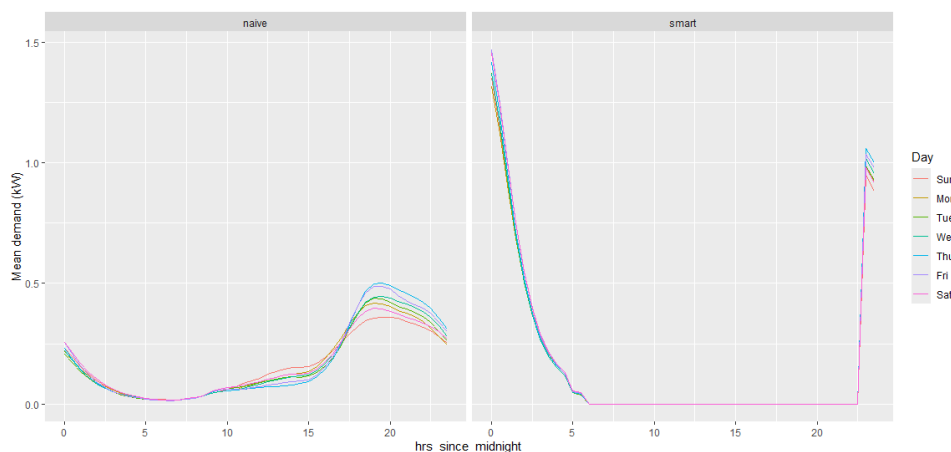


Figure 28: Average daily EV profiles across all EVs in the Electric Nation dataset.

One further processing step was required for both time series to align the timestamps with those in the CLNR demand data for summation to total demand per time stamp — since the CLNR data spans 2011–2013 and the Electric Nation EV data spans 2017. EV plug-in times and total charging session demands depend fairly strongly on time features such as time of day, day of week and time of year, but much less on weather conditions. As such, demands are relabelled in year-long chunks, matching to the same day of week and time of day. Seasonality is preserved while allowing variation in the exact shifting by matching day of year within a tolerance of 30 days. From this method, three year-long chunks are created and concatenated for each EV, matching the complete set of time points in the CLNR demand dataset. This resulted in a total of 583 possible EV time series to sample from. As in the CLNR household demand time series, missing values were filled with the mean for the same time of day and month of the year.

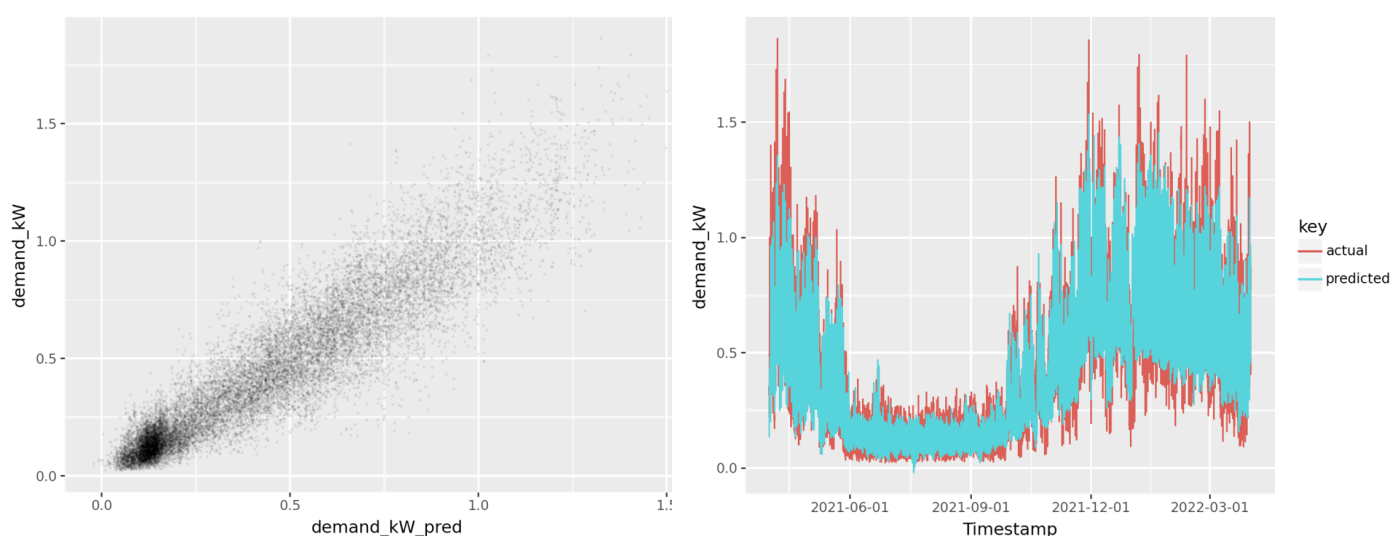
Processing of heat pump time series

As a technology that has not yet seen widespread adoption (but whose uptake is expected to increase in future), the set of available datasets on heat pump demands is limited. The electrification of heat demonstration project [45] published one year of heat pump demands for 742 households with a heat pump, across three regions of the UK (the North East and South East of England, and the South East of Scotland). The study specifically aimed to recruit participants across a range of building types and ages. The set of heat pumps was filtered for data quality and completeness, first dropping any heat pumps that were identified in the original project as having poor data quality (in terms of length or faulty data) and then also excluding any remaining heat pumps with more than 25% of half-hourly timestamps missing. This left a set of 73 heat pump time series.

Heat pump demands are expected to be related to not only weather (as heating demand increases in colder temperatures), but also housing characteristics such as size and insulation levels, and customer behaviour with factors like occupancy and preferred indoor temperature also influencing demand. These factors are expected to vary from LV network to LV network, potentially having a significant effect on the per-household heat pump demand that could be expected on different LV networks. Categorising the heat pump demand data by housing type and sampling only from the corresponding housing type would be possible, but would severely limit the set of possible heat pump demands available for each location so was not undertaken in this work. This means the results will not include the true variability in heat pump demands by location, which could exacerbate network demands and therefore cable damage in locations where all households are likely to have higher than average heat pump

demands.

As for the EV demand time series, some processing of heat pump demand time series was necessary to match the time points in the CLNR household demand dataset (2011–2013) as the original heat pump data was collected between 2021 and 2023, and across different date ranges for each individual heat pump time series. Previous work [46] has demonstrated the fitting of a model for the weather and time dependence of nationally aggregated heat pump demands that was successfully used to rescale time series to new time points. Figure 29 shows the relationship between actual and predicted heat pump demand for this model. There is a relatively strong relationship between predicted and actual demands, and the predicted demands capture the seasonal variability in heat pump demand. These features allow a weather rescaling method based on model predictions, but still including randomness from measured demands, to be implemented. This approach to time series rescaling was originally proposed for rescaling of national demand [47], but had been extended to this use case.



a: Average per-unit demands across whole dataset, actual vs predicted.

b: Time series of average demands.

Figure 29: Fitted model is able to predict collective heat pump demands, including seasonal patterns.

Applying the same model of weather and time dependence to individual heat pumps yields the relationship between predicted and actual demands in Figure 30. The ability of weather and time features to explain the demand of an individual heat pump is clearly much weaker than that of an aggregated group of heat pumps. There are also a proportion of times where actual demand is zero, i.e. the heat pump is not running, which is not seen in the aggregated data: the model is poor at predicting this. This is to be expected, as this switching will depend on much more than purely the weather: it is also a function of the users' preferred indoor temperature, and the building physics (size, insulation etc.), which are not known. Adding an extra layer with a binomial model for the on/off status of the heat pump was considered to include this behaviour, but was discounted for two reasons: the on/off switching has a much weaker relationship to explanatory variables (there is an element of randomness) so the skill of this model is expected to be limited; and secondly, the weather rescaling methodology assumes all points in the time series represent the same behaviour and relationships to weather. Any times where actual demands are zero will not produce physically meaningful residuals, making the previously employed weather rescaling method unsuitable. Therefore, weather rescaling was not employed for heat pump data scaling.

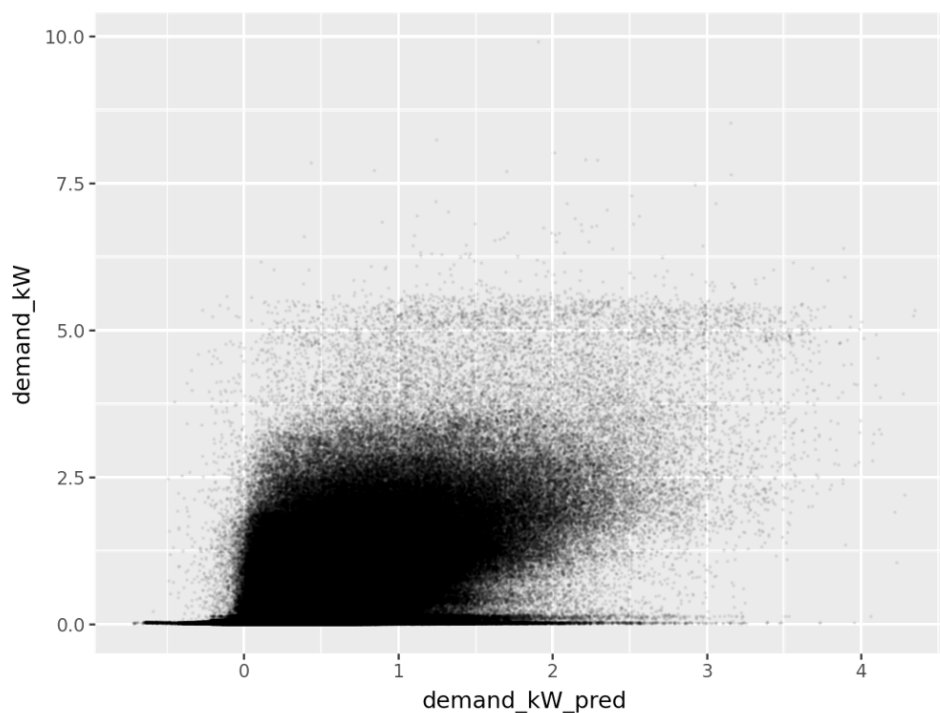


Figure 30: Predicted vs actual demands for an individual heat pump.

In the absence of other options, the timestamps of the heat pump data have been adjusted in the same way as for the EV time series: so any seasonality and variations in demand with time features will be preserved, but for a given time point, the resulting heat pump demand is not based on the same underlying weather as the household demands. This could reduce the extremes seen in total demand, as cold weather would simultaneously increase household and heat pump demands, but this effect is expected to be relatively small. The average demand profile seen across all heat pumps is shown in Figure 31. While there is an evening peak in heat pump demand, the largest demand is seen in the early morning when houses are heated after an overnight period of lower heating demand, presumably because occupants have programmed their heating systems to a lower temperature setpoint overnight, or turned their heating off. The magnitude of the evening peak in heating demand is larger than that of the domestic appliance demand evening peak, suggesting a house with a heat pump has the potential to more than double their evening peak compared to a house without a heat pump, with an even greater increase in morning demand.

In summary, there are currently very few data sources for real heat pump demands. Insight into the possible behaviour of individual, or small aggregations of, heat pumps is a knowledge gap that is particularly pertinent for DNOs.

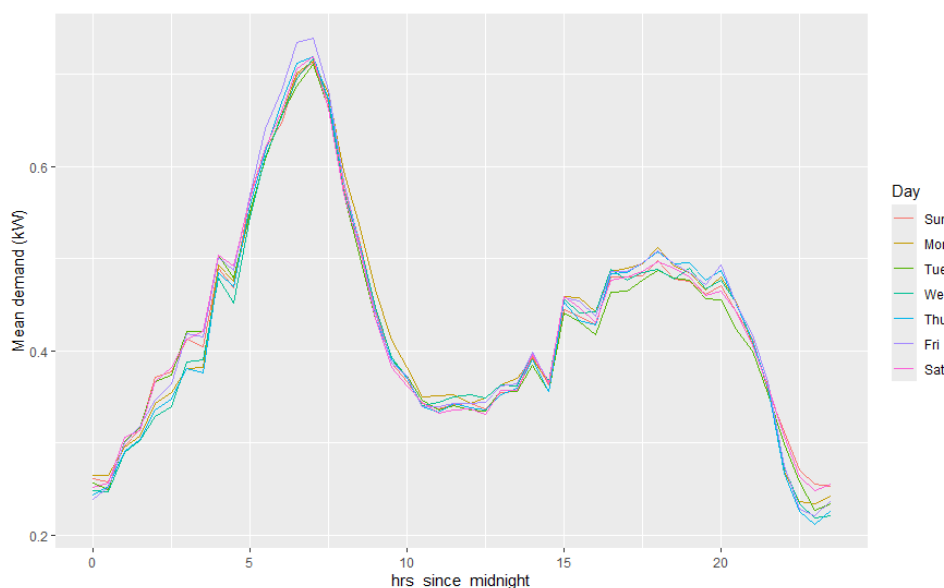


Figure 31: Average daily demand profiles across all heat pumps.

4.1.7.3 Demand summary statistics

While the full half-hourly time series of total demand is passed to the next step in the simulation process for simulation of cable currents, we also calculate some summary metrics. These are used as inputs to the statistical models built in Chapter 6. Total annual energy per year is calculated; this is assumed to be known for all customers and so is used as an input to the statistical models. This is calculated and retained for both the entire set of downstream customers, and the subset of smart meter demands. Quantiles of demand matching the probability levels modelled (q95, q98 and q90 daily range) are also calculated and retained for the total smart meter demand series - since the full half-hourly time series of demand is known for this subset of customers, this information would also be available as a potential statistical model input. Again, these are calculated per year as the statistical models are set up to predict per-year values, culminating in per-year damage predictions.

4.1.8 Step 8: Emulator

Due to the computational burden of conducting load flow calculations for such a large number — billions — of time points, a tailored, fast, machine learning model was developed to replace full load flow calculations in the simulation process. Separate emulator models were built for every cable on every network, capturing the individual relationship between demands and current in the target cable for that particular topology. Section 5.2 describes the full details of the emulator models.

Summary statistics of current are saved out from the current time series that are the output of the emulators: again, per-year q95, q98 of half-hourly current and q90 of the daily current time series are recorded as these are needed as target variables for training the statistical model for current, even though current is not expected to be known in practice.

4.1.9 Steps 9 and 10: Physical simulation of joint temperature and damage

The physical models set up to take input currents and properties of all cables around a joint, and return joint temperature and damage, are described in detail in Chapter 3. The dynamic temperature calculation upsamples the temperature time series to return a higher resolution time series of joint peak temperature. From this, the maximum observed temperature across each half-hour is retained, and the quantiles of this are taken. Summarising back to half-hourly values before taking quantiles in this way ensures what the quantile values represent, in terms of number of half-hours per year over the quantile threshold, is consistent with the quantiles calculated for current. Similarly, q95 and q98 of half-hourly peak temperature for each year in the data, and q90 of daily peak temperature range per year, are recorded.

The physical model for damage returns a single value of accumulated damage over the entire supplied time series. In order to return per-year values for training the statistical model, each calendar year of current data was supplied to the temperature and damage models separately, and the output value of damage for that year was recorded.

4.2 Simulation assumptions

Several assumptions have been made in order to produce simulations with as realistic and accurate results as possible, while maintaining computational tractability. Some assumptions and modelling judgements have been necessary due to the lack of data, or reasonable parameter values needed for the modelling have been assumed.

- 1. Demand time series missing values:** In the household demand, EV and heat pump half-hourly demand time series, missing values are filled with the mean for the corresponding time of day and month of year, from the rest of the data for that location. Filling missing values is necessary since the temperature and damage calculations require complete time series. This method of filling missing values ensures the overall mean of the time series isn't affected and daily and seasonal patterns are preserved.
- 2. Heat pump demands are sampled randomly, not accounting for different housing types in different areas:** The limited size of the dataset of heat pump demands meant further subsetting by housing type would leave too small a sample size for each subset. However, exploring the availability of a larger set of heat pump demands and considering accounting for housing type in the sampling of heat pump demands is a recommended future development of the model.
- 3. Heat pump and EV penetrations are assumed to be independent:** In reality there is likely some correlation between areas with high EV uptake and high heat pump uptake. However, assuming independence in these two variables ensures coverage of all possible LCT uptake scenarios in the simulation set.
- 4. EV and heat pump uptake in households with smart meters:** The EV and heat pump penetrations for household with smart meter data was assumed to be the same as the wider feeder LCT uptakes.
- 5. Not all joints were modelled:** Only joints with at least two physically different non-service cables, with at least one customer connected, were modelled. This ignores breach service connections as the resistance and therefore heating on these is expected to be much lower than main joints.
- 6. Missing cable properties were inferred:** Where values for cable properties were missing, these were inferred from the most common matching cable. For example, if only cable diameter was known, the cable type most common for that diameter was assumed.

7. **Any service points with no known addressed connected were assumed to have 0 load:** A check of these points confirmed they are either unused connections or most often unmetered connections such as street and traffic lighting; the demands from these connections is expected to be very low.
8. **Balanced phases:** Phase connectivity of addresses is not known, and full modelling of possible phase imbalances was beyond the scope of this project. However this was recognised as a potentially significant factor in increased joint temperatures and therefore gaining a better understanding of real network phase imbalances would be important to progress this aspect of the work.
9. **A power factor of 0.95 was assumed:** The types of loads that make up total household demand determine the power factor, and this value affects the results of the load flow emulator output currents. This best estimate value was based on relevant previous research [48].
10. **No industrial demands were modelled:** Industrial demands can be very variable depending on the type of industry, so modelling of typical industrial demands is not straightforward. Household demands were expected to dominate at LV.
11. **Some input datasets are older than others:** The EV and heat pump demand datasets include data from as early as 2011, and the EV data extends only to 2017. In the absence of more recent data, it is assumed that these datasets are a suitable basis for the model and for postulating future demand profiles.

5 Load flow modelling from GIS

The process of modelling LV networks is split into two main processes: first extracting the network topology into a useable form, and secondly constructing trained emulators that can be used to predict cable currents at any given point on the network from simulated demands.

5.1 Network topology from GIS data

This section introduces the GIS processing of network data, focussing on network assets within the ENW distribution area. This project focussed on the LV networks (below secondary transformers) that are the last stage of the distribution network which supplies customers. The current through (and therefore damage to) each of the cables on a connected LV feeder depends on all the loads on the feeder and the topology of the feeder. Therefore, it is important to model the LV feeder as a whole and simulate demands on realistic (or indeed, real) feeder topologies. Constructing networks from real GIS data for ENW's area ensured the resulting simulated data was representative of real network topologies across the area of interest. A range of localised demand conditions and load distributions were accurately captured by constructing models for multiple locations across the whole ENW area, with all physical joints modelled. Since each LV feeder serves a specific set of distinct customers with a different set of cables and topology, analysing a set of feeders with a range of cable types captured the variation in feeder configurations and cable loadings. This allowed representative modelling of how customer demands contribute to cable degradation across the whole network.

To model the impact of demand at specific locations, attributes such as cable types, diameters, and the number of customers served downstream must be known, as well as the topology of the network; all these factors combine with demand to affect the currents through each cable. To this end, GIS data was processed to identify the set of cables connected to a given transformer and to create a network representing the electrical connectivity from the transformer all the way downstream to the customers being served. Given the scale of the network (over 19,900 ground-mounted transformers) it would have been computationally infeasible to simulate all of them. Instead, a random sample of 111 transformers was taken, ensuring the sample size is large enough to include variation seen across the complete set in terms of geographic location, size of network and urban/rural locations. Simulations of demands, temperature and damage for this sample of transformers provided the training set for later statistical models.

Since each transformer can have multiple feeders, each with many cables, this still resulted in a significant dataset for simulation and later statistical model training. Figure 32 shows the distribution of the subset of 111 transformers, randomly chosen to represent the overall population. Figure 33 compares the distribution of transformer ratings for the whole network population with that of the subset of 500 processed transformers, showing that the distribution of the ratings of the sampled transformers is generally representative of the population.

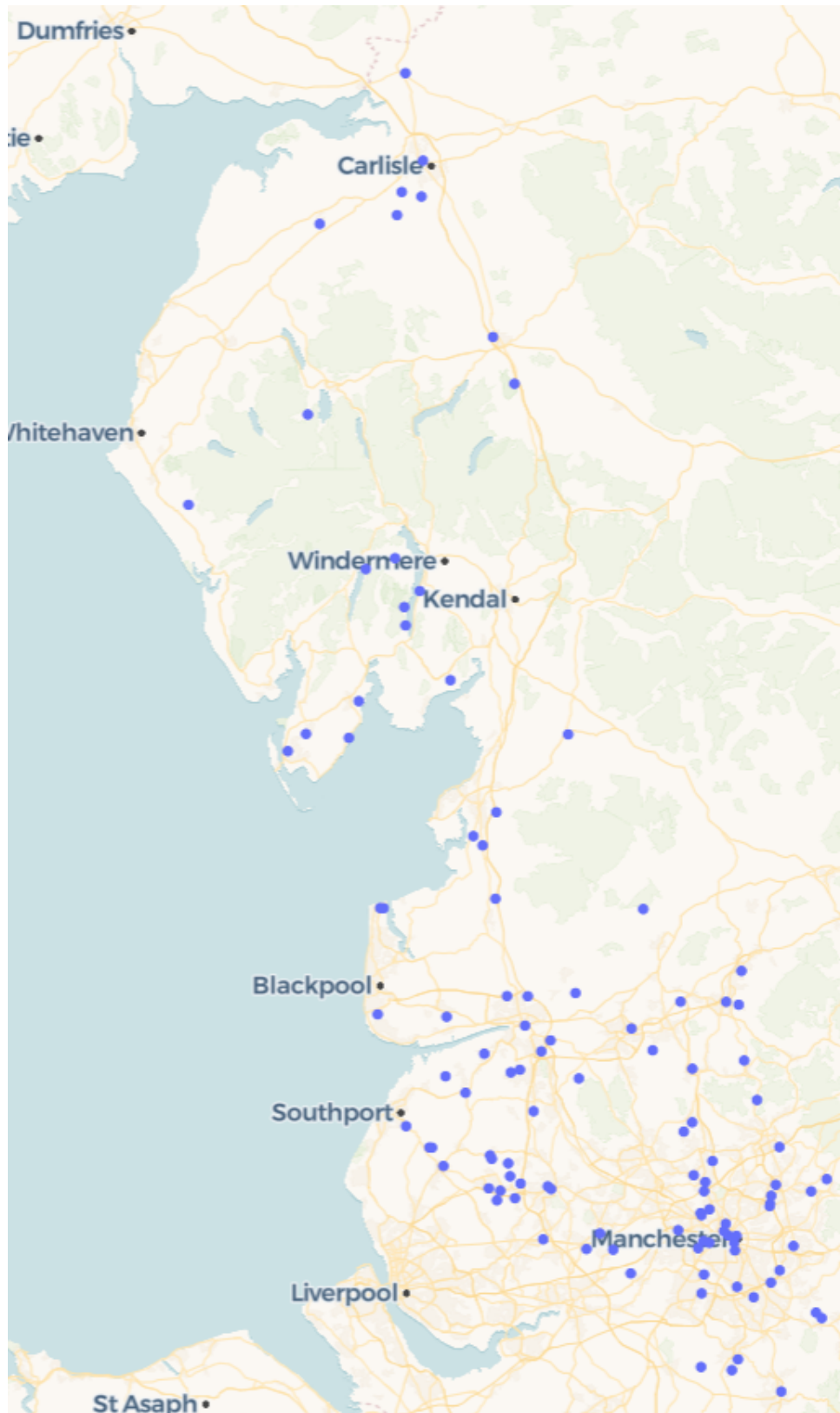


Figure 32: Geographic distribution of subset of selected transformers.

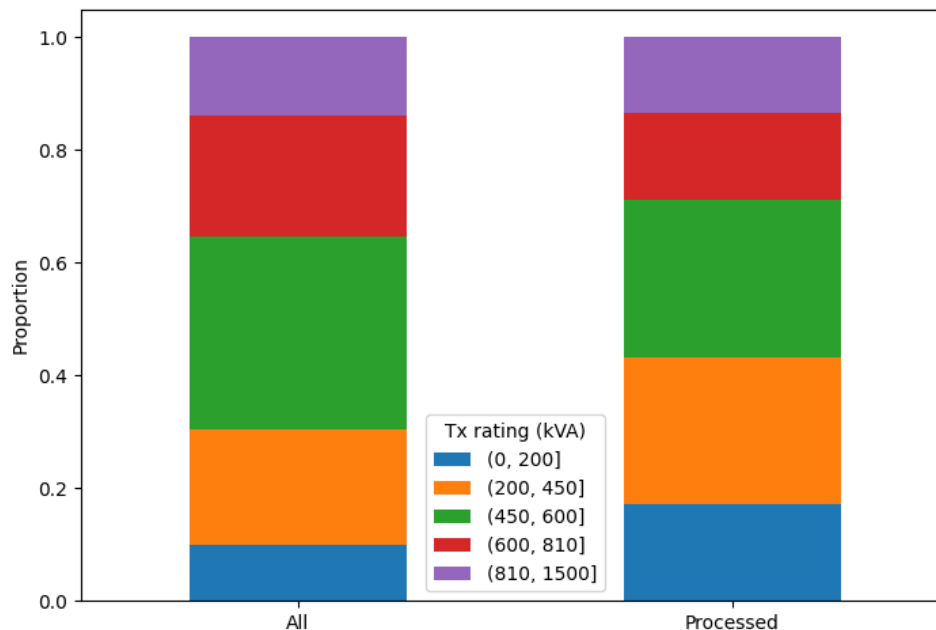


Figure 33: Comparison of the distribution of transformer ratings for the sampled subset to the whole network.

For each selected transformer, the connectivity of all feeders from the transformer are processed by detecting the intersection of joints and cables. To efficiently track the path within the network, a bounding box is used, allowing us to search for all cables within a certain distance of the transformer without needing to load the entire GIS dataset each time. As the path is tracked within this defined area, LV cables and service cables encountered along the way are recorded. This connectivity, along with the cable properties and connection points of addresses, allowed for the assigning of demands for simulations of possible load flows in each cable on the feeder. In this way, we can derive realistic time series of currents in any given cable, based on sampled loads.

5.1.1 GIS processing methodology

The GIS processing contains several individual steps, summarised in Figure 34 and described below:

- ▶ **Processing LV GIS data:** This step detected all the intersections of cables in the network within the specified area around the distribution transformer, starting from the top of each feeder. For the purposes of mapping connectivity, cables were split (and extra nodes created) at all cable intersection points to ensure accurate linking of joints and cables. Original cable IDs were retained to ensure knowledge of physical cables present, and new unique IDs were assigned to each split section for tracking purposes.
- ▶ **Identifying connectivity:** Starting from the transformer and working down each feeder, joints at the end points of each cable were found, and the next cables connected to those joints were identified and added to the queue. The process was continued until the connectivity of all cables on each feeder were identified as a sequence of joints and cable connections.
- ▶ **Network Graph:** A graph was created to represent the feeder connectivity. Nodes represent joints and service terminations, while edges represent cables (both service and LV cables). The graph contains both the network connectivity and the physical attributes of each cable, through assigning attributes to the graph edges. An example network graph is shown in Figure 36. The graph form of the network allows calculation of

the cumulative number of customers or LCTs for each cable section up the feeder, classification of taper joints (where cable diameter changes), and determination of which demands are downstream of any particular cable. This is used for feature engineering in the emulator model training.

- **Network models:** The network graph from the previous step was used to create an electrical network in a format readable by power system simulation software, in this case, IPSA. This included creating components for the transformer and all cables on the feeders, with joints as busbars and cables as line components. Load components are also assigned at every service point with customers connected, with an initial nominal value and the actual values are updated based on sampled demands before any load flows are run. The electrical properties of all these assets were set in the power system simulation tool (IPSA), including transformer ratings and voltages, and cable resistances and reactances. Cable resistances and reactances were calculated from cable type, conductor area and length. Resistance and reactance per unit length were found from data sheets [49], matched to the corresponding cable type and conductor area. Where there was no direct match to conductor area, values were linearly interpolated from the two nearest cable sizes. Per-unit length values were then scaled by cable length derived from the GIS data to get cable resistance and reactances assigned in the network model. IPSA takes resistances and reactances as per-unit voltage:

$$\begin{aligned} R_{pu} &= R \times S_{base} / V_{base}^2 \\ X_{pu} &= X \times S_{base} / V_{base}^2 \end{aligned} \quad (25)$$

where S_{base} is 100 MVA: this is the conventional value used in Interactive Power Supply Analysis (IPSA), although it does not matter what value is used as long as it is supplied along with the pu values. V_{base} is set to the transformer outfeed voltage from the GIS data, so the value 0.415 kV is used. Again, V_{base} is supplied along with the pu values to IPSA when setting up the load flow calculations.

These network models contain all information about the network needed to conduct load flows, except the sampled demands which are added separately as they vary for each simulation.

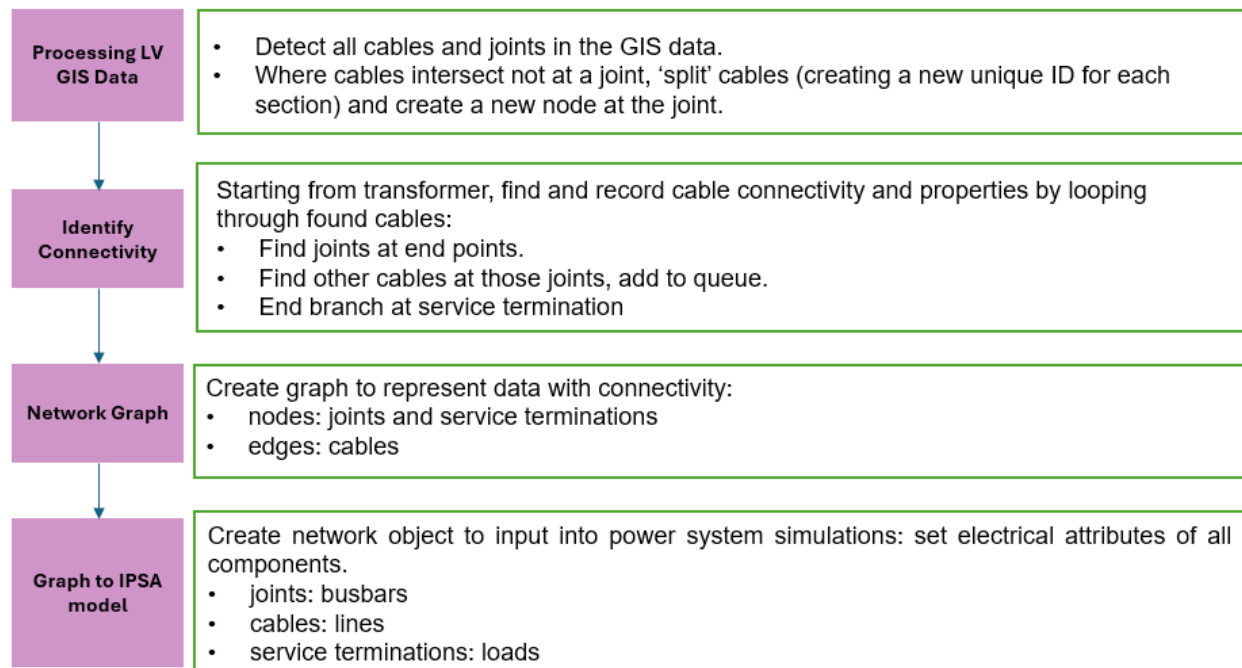


Figure 34: Flowchart of steps for GIS processing and generation of network models for load flows.



Figure 35: GIS data for example feeder.

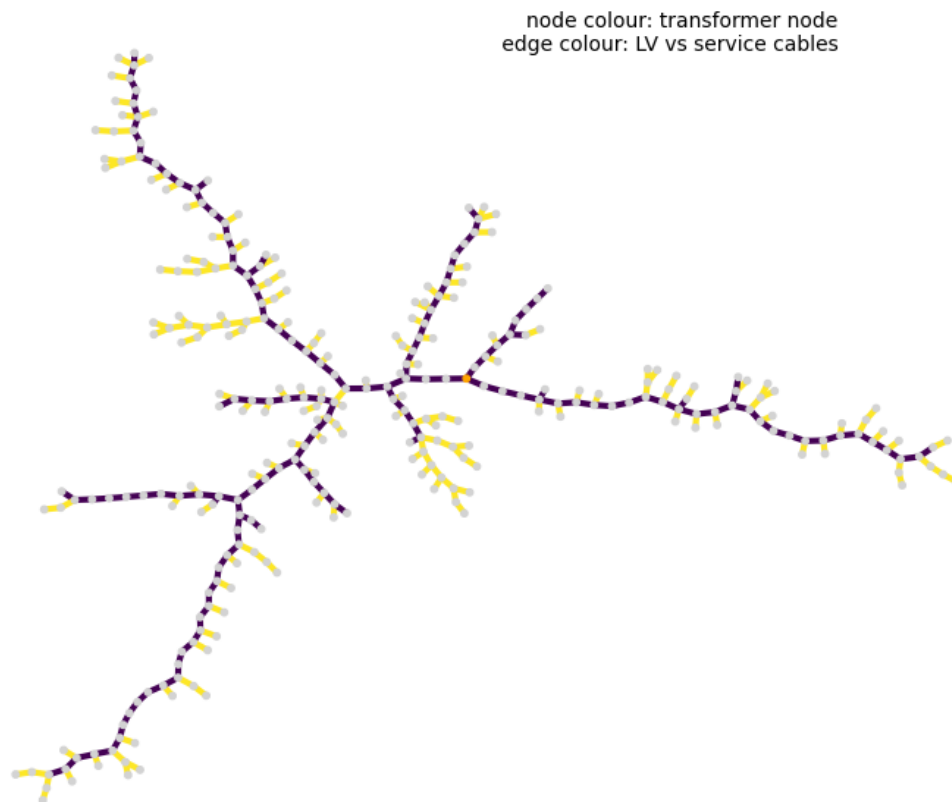


Figure 36: Connectivity for example feeder. Orange node=transformer; purple edges=LV cables; yellow edges=service cables.

5.2 Load flow analysis

Load flow modelling was used to realistically model the current flowing at any point in a given LV network based on the network topology, physical cable properties and demands sampled to points with load connected on the network.

The computation time of load flow analysis quickly becomes a bottleneck due to the large volume of samples and network models required for the statistical analysis of thermal degradation across ENW's LV network. Leveraging IPSA enhances performance to support these analyses, but trade-offs exist between load flow methods, particularly in terms of accuracy, convergence speed, and computational complexity. IPSA's proprietary engine dynamically balances these factors to optimise efficiency. For a range of LV networks containing between 50 and 500 busbars, a cluster of load flow calculations covering three years of half-hourly data takes approximately 1.5 to 2.5 minutes to compute, respectively. Despite this performance, generation of current time series for many network topologies, cables, LCT penetrations and demand time series, required running thousands of sampled multi-year time series for each cable across various high-density LV networks, making traditional load flow methods time-consuming. To overcome this computational challenge, an active learning approach has been introduced as a proxy model for load flows. This method reduces the number of explicit load flow calculations required by several orders of magnitude, making the large-scale statistical analysis based on MC sampling feasible. By combining machine learning with traditional power system simulations, active learning enables rapid and scalable assessment of network performance under diverse operating conditions.

Active learning is a machine learning technique that selectively queries a subset of data points to minimise computational overhead while maintaining high accuracy. Active learning is used to efficiently build a proxy model for load flows, called an emulator. Instead of running full load flow simulations for every possible condition, an emulator model is trained using a strategically chosen subset of initial training data that provides broad coverage of input variations. The model is iteratively refined through active learning, selecting new data points based on uncertainty metrics and only running load flow simulations when predictions from the trained emulator lack sufficient confidence. The emulator not only estimates the results of IPSA load flows but also quantifies uncertainty in its predictions. When uncertainty is low, predicted values are assumed to be sufficiently accurate. When uncertainty is high, new load flow simulations are run, and the model is updated dynamically. This ensures computational efficiency while maintaining model accuracy. The modAL python package provides a flexible framework for building active learning models [50] with Figure 37 summarising the type of high-level structure used.

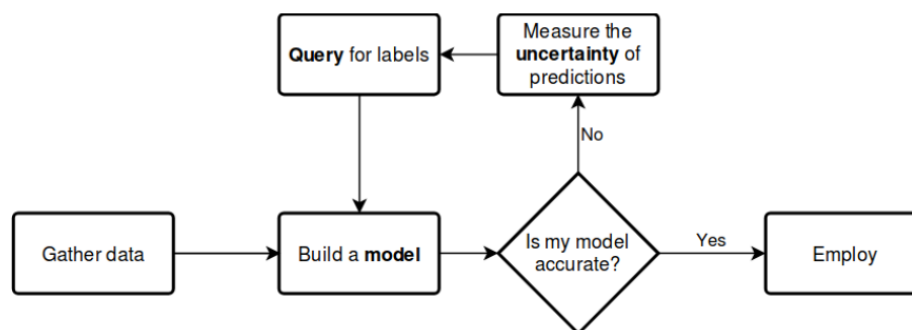


Figure 37: High-level structure of an active learning model. Taken from the modAL documentation.

5.2.1 Advantages of using emulators

Emulators are only appropriate for certain applications where there is a well defined relationship between the input features and target value, where large volumes of predictions are needed and where there is a significant time cost to generating these values empirically (for example through computationally involved simulation or calculation, or manual labelling). If appropriate for the application, emulators offer several key advantages:

1. **Overcoming computational bottlenecks:** Traditional load flow simulations are time-intensive, especially when analysing thousands of multi-year time series for each cable. Active learning minimises direct load flow simulations, instead predicting results from a fast machine learning model which is trained on a strategically selected subset of training data.
2. **Optimised trade-offs between accuracy and efficiency:** The active learning framework balances computational complexity, convergence speed, and predictive accuracy. By leveraging feature engineering and an appropriate choice of model forms for the application, the method ensures that results remain reliable while significantly reducing computation time.
3. **Generalisation:** The implementation is general and applicable to a diverse range network topologies, as regression models are fitted to make predictions per network component with separate emulators trained for each unique network topology. This compartmentalised approach allows for scalable emulation of IPSA load flows across diverse network topologies.

4. **Robust uncertainty quantification:** The model does not just predict load flow results but also quantifies the uncertainty in these predictions. If predictions fall within an acceptable confidence range, they are used directly. Otherwise, IPSA load flow calculations are triggered to enhance model training and ensure confidence in the returned results.
5. **Tracking of model performance:** The performance metrics (R^2 and MAE) are recorded and the performance of all models is evaluated out-of-sample on these metrics. This ensures accuracy in the returned predictions is maintained.

5.2.2 Emulator structure

The active learner consists of two models, one for predicting current and one for the uncertainty on those predictions.

5.2.2.1 Current model

The main predictions of current use piecewise linear model to predict the mean current, ultimately replacing the load flow calculation. This has the following properties:

- ▶ Allows different linear relationships across the range of input feature values, maintaining a balance between simplicity and flexibility of shape fitted
- ▶ Inclusion of a regularisation penalty term shrinks the contribution of the least informative inputs: this allows automatic feature selection and de-selection of demands that are uninformative for the target cable. Elastic net regularisation² is used, with automatic cross-validation³ to find the optimal hyperparameters for this.
- ▶ Feature engineering: only demands downstream of the target cable are supplied as input features, as well as the total downstream demand. This reduces the complexity of the model, enabling more efficient fitting and better performance from the model regularisation since the least informative demands are already excluded.

Demands are labelled as 'downstream' of a given target cable if there is a path between the demand and the transformer that passes through the target cable. This is determined using the graph representation of the network and is used to produce a dictionary with a key for every target cable, and where the corresponding value contains a list of the demands which are downstream of that cable. The function `all_simple_edge_paths` in the `networkx` package is used to list all edges on all paths between two nodes. Figure 38 demonstrates the expected result for a simple case.

Grouping of demands to reduce the number of features supplied to this model was also tested. This used the graph representation of the network to group all demands within a specified threshold impedance of each other, since electrically 'close' demands will act similarly to one aggregated load. However, this additional feature engineering was not found to add additional skill to the emulator and that supplying all individual downstream demands (including regularisation) was sufficient to produce skilful predictions.

² For technical detail, see Section B.2

³ see Section B.3

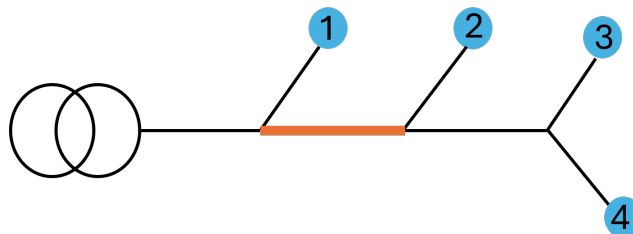


Figure 38: Simple example network. Here loads 2, 3 and 4 are downstream of the orange cable.

5.2.2.2 Uncertainty model

The uncertainty model makes predictions of the q16 and q84 quantiles which are combined to give a prediction of the central 68% range (i.e. roughly one standard deviation above and below the central value). The exact probability interval this covers is not important, but it gives a consistent estimate of prediction uncertainty that can be used to compare the uncertainty of the model's predictions for all time points. LightGBM models [51] were used to generate these predictions: this is a fast implementation of gradient-boosted decision trees, which inherently can accommodate nonlinear relationships between the input and target variables, interactions between input variables, and feature selection and regularisation through control of the depth of the trees. Predictions of uncertainty from this model are used to determine the most uncertain points in the prediction set; if the uncertainty on these points is unacceptably high, they are passed to run the actual load flow on. This reduces the uncertainty on these points, as they are now included in the training dataset with 'actual' values known from the load flow. Retraining the emulator in this way improves its skill in the most effective manner, minimising the number of time points the full load flow calculations are run for.

The emulator is implemented through a structured process, which builds a model framework to approximate IPSA load flow results while selectively refining predictions based on uncertainty metrics. This is split into two processes: the initial building and training of the emulator, and then prediction from the trained emulator. The overall process is outlined in Figure 39 below.

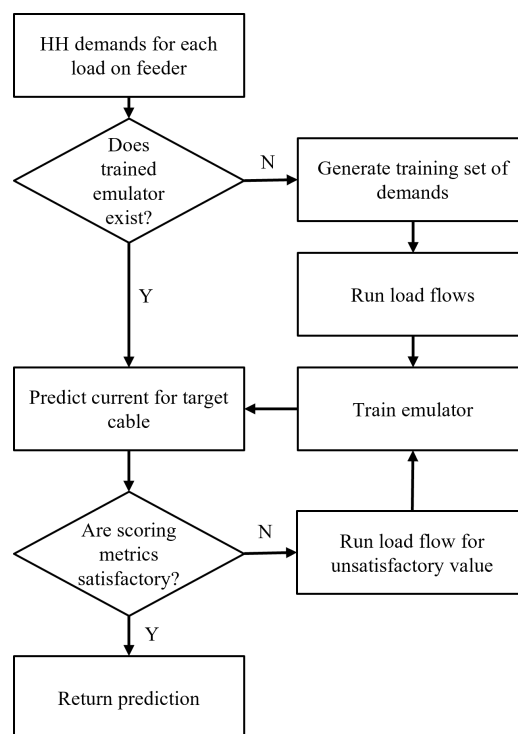


Figure 39: Flowchart of the general process for emulator training and prediction.

5.2.3 Building emulator framework

Before it can be used to make predictions, an emulator must be set up and trained for each target cable. The key steps to produce the emulator model are set out in Figure 40, with the individual steps in the process described in more detail below. This is an iterative process, where load flows are run for more samples and the emulator is retrained until satisfactory performance is achieved. Only then can it be used in the wider simulation process to make predictions

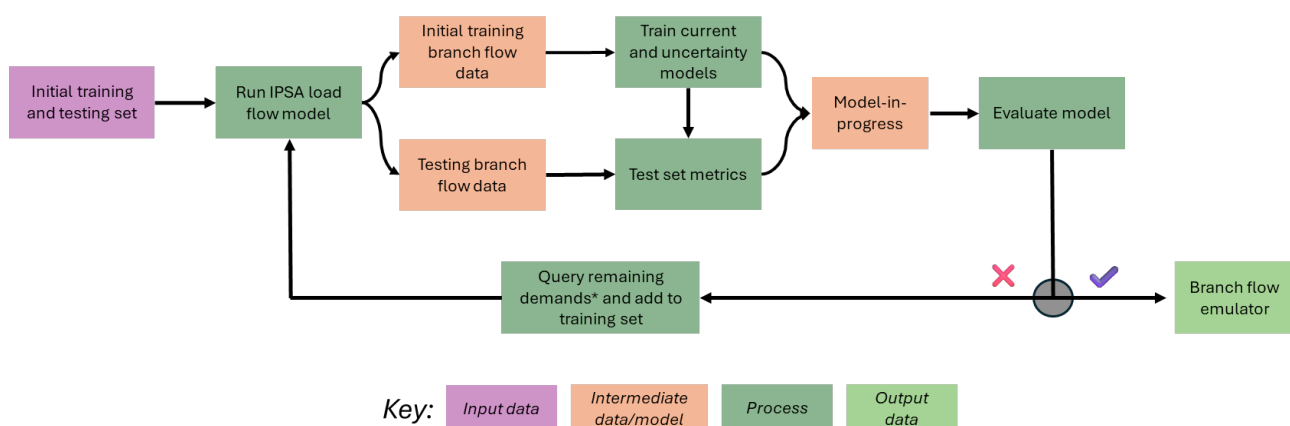


Figure 40: Flowchart of initial emulator training process.

► Generation of initial training and testing set

A technique called Latin Hypercube Sampling (LHS) was originally used to sample the initial set of demands

that will be passed to a load flow and used as the initial training set for the emulator. The LHS technique creates a stratified sample set across many dimensions, ensuring each input variable (demand in the network) is sampled evenly across its entire range. LHS minimises gaps and clustering, improving learning by ensuring all regions of the feature space are well represented. The LHS set was generated by specifying the limits (maximum) for each demand, and a correlation coefficient between the demands, which was taken from measuring correlations between sample demand time series. However, while the correlation between demands is specified, this does not include the full dependence between demands seen in the full sampled time series. Perhaps because of this, and the condition of equally spaced values in a high-dimensional space, the LHS sampled inputs may not be entirely realistic and this method of sampling resulted in a large number of nonconvergent load flows for some networks. Much better convergence was seen for a set of time points based on real sampled demands, suggesting the nonconvergent time points were not representing likely real demand scenarios. Due to the additional computational burden of running a larger set of load flows and then needing to discard the proportion that did not converge, we chose not to use LHS and instead sampled the initial training and testing data from the set of sampled demands that the simulation process was set up to predict for, X_s . This ensures the training data inputs are realistic. Additionally, to ensure as even a distribution over all feasible demand values as possible, the large set of all demands, X_s , are ordered by total demand before a stratified sample is taken. This ensures the training data includes the full range of possible total demands on the network. In addition, the single sample with the highest demand for each individual load is included in the training data to ensure the full range of possible demands at each load are represented. While the number of customers on a given network is fixed for all simulations, runs with a larger number of LCTs were expected to have larger maximum demands. To ensure the training data (used for all simulations for the same network) covered the range of all possible demands seen under any simulation, the simulation with the largest number of LCTs was run first: the first simulation is the one that generates the emulator training data set (subsequent simulations load and use the pre-trained emulator models from this first run). From testing for a range of network sizes, an initial set of 1000 time points that were split into training and testing sets was deemed sufficient for good emulator performance.

- ▶ **Running load flows:** Once a representative set of input demands has been selected, these are passed to a load flow model: the network form described in Section 5.1 includes all physical properties of the network needed for the load flow calculations and forms the base of this model. For every sample in the training set, demand values are set for each load on the network and a load flow run. The resulting currents are extracted for all target cables on the network. This results in a table with $n_{demands}$ columns and $n_{samples}$ rows as the input to the load flow, and a table with n_{cables} columns by $n_{samples}$ rows as the output. This initial dataset of inputs and outputs are generated separately for each physical network and saved out to be used to train all the emulators for that network.
- ▶ **Splitting samples into training and testing data:** It is important to hold back a set of samples where both the inputs (demands) and outputs (currents) are known, to test and verify the performance of the emulator. The set of 1000 samples with known demands and currents is split randomly to produce a training set (with 2/3 of the samples), and a testing set (with the remaining 1/3 of samples). The test set is only used to calculate performance metrics for the emulator.
- ▶ **Initial training of the emulator:** Separate emulators are trained for every target cable in the network. The emulator is trained from scratch only the first time a simulation is run for that network; once a trained emulator exists, it is saved out and can be loaded in for making predictions for future simulations on that network. To train the emulator, the input demands are supplied as features and the `modAL` class structure trains both the

central GAM and the uncertainty (quantile lightGBM) model from these features in one call. Before model fitting, inputs are scaled to ensure a consistent range across all features, and this scaling is retained and also applied to all inputs when predictions are made from the model.

- **Evaluation metrics:** The performance of the trained model is then evaluated. Two metrics are used to assess the accuracy of the model's central predictions of current: R^2 and MAE. Here the following definition is used:

$$R^2 = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2} \quad (26)$$

where y are the actual values of current from the load flows, $\hat{y}$ are the predictions of current from the emulator, and $\bar{y}$ is the mean actual current. This metric can be thought of as a measure of the proportion of variance in the real data that is explained by the predicted model. An R^2 value of 1 would represent perfect predictions. MAE measures the magnitude of the average prediction error, giving an indication on the scale of the target variable (in this case, current in A) of the additional error introduced by using the emulator instead of running load flows for all time points. This is defined as:

$$\text{MAE} = \frac{1}{n} \sum |y - \hat{y}| \quad (27)$$

where n is the number of samples evaluated.

- **Evaluate model:** Thresholds are set for acceptable values of R^2 and MAE, and the model performance is evaluated against these thresholds. If performance on both metrics is satisfactory, the emulator is considered trained and ready to use for predictions. If not, active learning query strategies are employed to generate more samples of demands and currents, which are then used to retrain the emulator before its performance is re-evaluated. Two main query strategies may be used:
 - **Max-error query:** This query strategy requires a set of samples where the target (current) is already known (a 'retraining' set). From this set, the prediction error of the current emulator can be calculated, and the m samples with the highest error are selected to retrain the emulator. If several iterations of emulator retraining are expected, and/or the computational cost of generating new samples with known currents is low, this method can result in an emulator with greater skill than would be achieved by purely training on all available input-target pairs. It is essentially a way of picking the most informative points in the retraining set. However, there is an upfront computational cost in generating the retraining set for this query strategy, and a proportion of the samples will be run through a load flow, but not used to train the emulator, so it is less efficient.
 - **Max-uncertainty query:** This query strategy takes a set of samples where the target (current) is not already known, and makes predictions of the *uncertainty* for each of those samples. Samples with higher uncertainty are likely to be those that are less similar to points in the existing training dataset, and where the emulator has lower skill. The m most uncertain samples are selected, and these inputs are run through a load flow to generate the corresponding known currents. These points are then added to the emulator training set, and the emulator is retrained.

Given the higher computational cost of performing the max-error query, the max-uncertainty query method was selected. Thresholds of $R^2 = 0.98$ and MAE= 1A are suggested to ensure reasonably high performance across all networks. For speed of processing and to understand the range of emulator performance across different networks, a more lenient threshold could be used, and the resulting R^2 values used to weight the samples when they are used as training data in the statistical models (detailed in Chapter 6).

For networks that do not meet the convergence criterion after a certain number of iterations of retraining, there are two possibilities: either they are still continuing to improve and will eventually converge to a satisfactory level of performance, or they have already converged but the level of performance is not satisfactory. It is important to be able to distinguish between these two cases, as in the first case there may be benefit to continuing to increase the training dataset, while in the second case this will not result in model improvement. Several measures of convergence were considered; Bloodgood [52] suggest a stopping criterion should be general (compatible with any emulator model and retraining batch size, etc); aggressive (favour stopping earlier and reducing computational time); and stable (i.e. fit for purpose across different applications and in our case different networks). They review the range of stopping criterion in the literature and suggest stabilisation of the uncertainty metric over a very large set of predictions is a suitable criteria. We also calculated rolling median gradient of R^2 to explicitly track the change in performance of the emulator over retraining iterations, although, this metric is more sensitive to varying scale for different networks and cables.

5.2.4 Predicting responses

Once an emulator has been trained for the network and target cable being modelled, it can be used to generate predictions for a large number of new load flow scenarios very quickly and efficiently. The prediction process also includes a feedback loop, assessing the accuracy and confidence of the predictions and reverting to running a load flow for any cases where the emulator prediction are not deemed satisfactory by these metrics. The process outlined below sets out the steps involved in the prediction and conditional retraining of a dataset for prediction.

If a sample lies outside the min/max range of input features, it should be run through a load flow as the underlying emulator models are known to have poor performance when predicting outside the range of the training data. However, the strategy of basing the initial emulator training on the highest-demand simulation for that network, and forming the emulator training data from a stratified sample of the possible demands for that simulation plus the single highest-demand cases for each individual load, ensures the highest possible demand cases for that network are always included in the training data. This ensures the emulators are suitable for use with all subsequent demand inputs.

There is no strict rule implemented in the emulator to predict only positive values; because of this, a small proportion of predictions for low currents and/or exceptionally uncertain points may return negative values. Negative currents would affect the subsequent temperature and damage calculations, so these must be corrected. The most rigorous approach would be to pass these time points, and any points with an unacceptably high prediction uncertainty, to the load flow model to get perfect predictions for these time points. Given the need to minimise the computational time of the process and since negative values occur rarely and for individual time points (not prolonged periods) as well as most of these values being only slightly negative and corresponding to low-demand time points, a faster approach of interpolation based on the predicted currents either side of the negative value was used.

5.2.5 Emulator performance

For the application of an emulator to LV network load flows, the relationship between input demands and resulting current is well defined. As the results are outputs from a load flow calculation, there is no noise in the training data and the same demands always result in the same predicted current. While the number of input demands can be quite high for larger networks, the relationship between input demand and target current is simple enough that we found very good emulator performance with low uncertainties from a relatively small number of training samples. The routine of iteratively retraining the emulator on the most uncertain samples was implemented in

model development, but this was found to give very little to no improvement in emulator performance and predicted uncertainty, with a large additional computational cost for running extra load flows for the highest-uncertainty samples. Because of this, simple performance thresholds on R^2 and MAE were employed and no retraining based on the most uncertain samples was used in the end model. The out-of-sample performance metrics were retained to ensure only emulators with very good performance were used.

Figures 41 and 42 show the distribution of MAE and R^2 for the set of emulators used for the 2000 unique cables modelled in the simulation process. These models have extremely good accuracy, with 96.6% of cables having an MAE models having an MAE below 0.25 A and 98.8% of cables having an R^2 of at least 0.99.

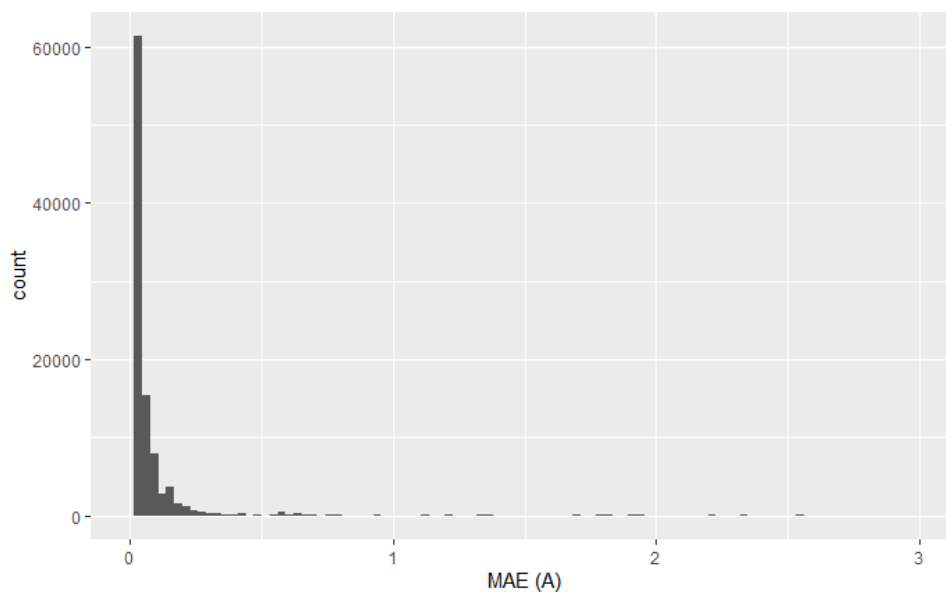


Figure 41: Distribution of emulator MAE across all modelled cables on all networks.

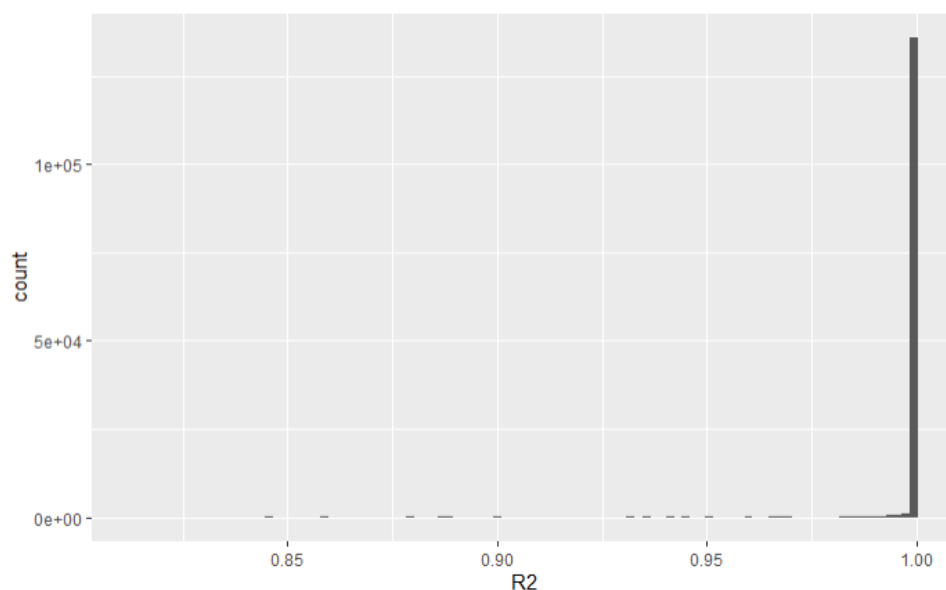


Figure 42: Distribution of emulator R^2 across all modelled cables on all networks.

5.2.6 Emulator assumptions

Some key assumptions were necessary to keep the complexity of the emulator manageable, and based on the constraints of the available data. Notably,

- ▶ Only demands are supplied as inputs to the emulator. Base voltage level, X:R ratio and cable capacitance are not explicitly included in the emulator, and are assumed relatively constant within each model. These factors are implicitly accounted for in the specific relationship fitted separately for each target cable on each different network.
- ▶ Only *downstream* demands are supplied as inputs to the emulator. While other demands on the network will also have an influence on the cable current, their effect is much smaller and the majority of the skill in the emulator's predictions comes from the input downstream loads. Including total upstream demand was found to have limited to no benefit, especially for more complex networks with multiple branches. Similarly, no benefit was found from grouping electrically close loads and supplying this smaller feature set.
- ▶ A constant power factor is used in load flows, set to 0.95 for all loads. This assumption could be relaxed by asserting individualised constant power factors for each substation based on historic substation monitoring data. However, while this would improve the accuracy of the modelling compared to today's network, this would not reflect future changes in power factor from increased LCT penetrations in the future. Power electronics in EV chargers can introduce reactive power effects, however modern chargers are designed with power factor correction. Heat pumps will typically operate with a power factor between 0.9 and 0.95 due to the compressor induction motors.
- ▶ Only balanced load flows are conducted. Investigation of the potential effect of phase imbalance was conducted (see Section 8.1), and this was seen to be a significant potential contributory factor to cable and joint heating - heating is proportional to the square of current, so current imbalances result in even larger heating imbalances. However, full modelling of phase demands, and running of a full unbalanced load flow was deemed infeasible based on the limited knowledge of customer phases at present. This is an important area for future exploration and development of the project and models.
- ▶ A protocol for resolving non-convergent samples is included in the load flow calculations. While the vast majority of samples converged, some samples did not; in these cases, a set of methods are attempted to return a converged load flow result. First, the number of load flow iterations is increased. If this is not successful, a warm start is attempted by gradually increasing the loading. If this still fails, locked transformer taps are unlocked, and finally if this also fails, the grid infeed voltage setpoint is increased. This protocol prioritises the resolution methods that are most likely to be successful with minimal changes to the original network.
- ▶ The full potential of active learning to effectively select the most informative samples for retraining has not been fully utilised, as the initial training was found to perform well across the set of networks. A smaller initial training set with some retraining through active learning could have produced equally skilled models with fewer training points; however, the differences between networks meant it was difficult to implement a fully automated stopping criterion that was valid for a large set of networks. Iterative retraining also added additional computational expense, as it requires selecting and running load flows for additional points for each target cable on the network rather than producing one training set that can be used to train emulators for all target cables.
- ▶ The active learner is trained to predict load flows in the supplied network only, i.e. the topology of the modelled

network comes from the GIS data and was assumed accurate. The whole load flow process sits within the wider sampling and physical simulation process, which is designed to create a dataset of results for plausible real networks. The complete accuracy of the exact topologies used is not absolutely essential in this, as long as the topology is realistic, known and consistent throughout the modelling process. This limitation does mean a separate, new emulator must be trained for every different network topology that is simulated and every target cable on that network, increasing the computational intensity of the process. However, the same set of emulators once trained can be used for multiple simulations of the same network.

6 Statistical modelling of current, temperature and damage

Physical simulation of many sampled scenarios on ENW's network has produced a dataset of possible realistic cable currents and joint temperatures and damages given input demands. However, this physical simulation process cannot feasibly be used to directly make predictions of damage for all joints in ENW's network for two main reasons. Firstly, the computational complexity is simply too large to run full simulations for all 19,000 LV networks (a subset of 111 were used in this project); and secondly, the physical simulation process assumes known half-hourly demands for all customers on the network, which is not available in practice. Instead, the set of physical simulations can be seen as a training dataset from which to fit statistical models which link simpler, known features of the network to predictions of current, temperature and damage.

6.0.1 Model target variables

The ultimate aim is to predict joint damage, so per-year joint damage is the final output of the statistical models. Modelling per-year values allows flexibility in making predictions for different years for different scenarios, allowing the models to answer various 'what if' questions. For example, predictions can be made for years in the future with different LCT penetrations, and trajectories of LCT uptake can give damage predictions for overall future pathways and scenarios. The uncertainty in age and historic damage of a cable can also be modelled by sampling different possible accumulations of per-year damage.

Working backwards from damage determines the most informative predictors to model through the chain of statistical models. The main drivers for damage are prolonged high temperatures and temperature cycles (as discussed in Chapter 3); therefore, we expect both high quantiles of temperature and a measure of temperature cycling to be informative predictors of damage. Models are trained to predict target temperature variables of q95 (the temperature threshold that is exceeded in 5% of half-hours in the year — 876 time periods) and q98 (exceeded 2% of the time, or 350 half-hours). The daily range in temperature was chosen as a measure of temperature cycling, and the 90th quantile of this (exceeded 36.5 days of the year on average) taken as the target variable for temperature range. A slightly less extreme probability level was used due to the lower data resolution this metric is based on, but was still found to have strong correlations to damage.

High quantiles of temperature are naturally linked to high quantiles of current, (and high quantiles of temperature cycles correlated with high quantiles of current cycles) since current directly influences temperature. The same quantiles are modelled for current, used as inputs into the temperature models.

6.0.2 Modelling of uncertainty

Uncertainty on unknowns is also naturally included in these models by explicitly modelling not just the deterministic central 'best estimate' of the target variable, but instead modelling the distribution of the target variable probabilistically. For example, the known feature of customer demand is total annual consumption, but the same annual consumption could have resulted from many different half-hourly profiles of demand that would have different effects on current, temperature and damage — so there is a *distribution* of damage values associated with a given annual customer demand, even for the same network. This variation in half-hourly demands (and its effect on current, temperature and damage) for the same total annual demand is contained within the training dataset through many examples of different sampled half-hourly demands and different network topologies.

Each single simulation in the training dataset is generated by simulating a full year of half-hourly demands, which are then translated through physical models into half-hourly currents and temperatures. The target variables of the statistical models, for example the 95th quantile, are then thresholds from the overall distribution of half-hourly values. Each simulation in the training dataset can be thought of as a single sample from the entire possible set of demand time series, resulting in summary statistics (q95, q98 and q90 range) of that sampled distribution. Just as the central limit theorem states that the distribution of sample means is normally distributed for a large number of samples, the normal distribution also provides a good fit for the target quantiles of current and temperature as demonstrated in Figure 43. This knowledge was used to pick appropriate modelling structures for the current and temperature models.

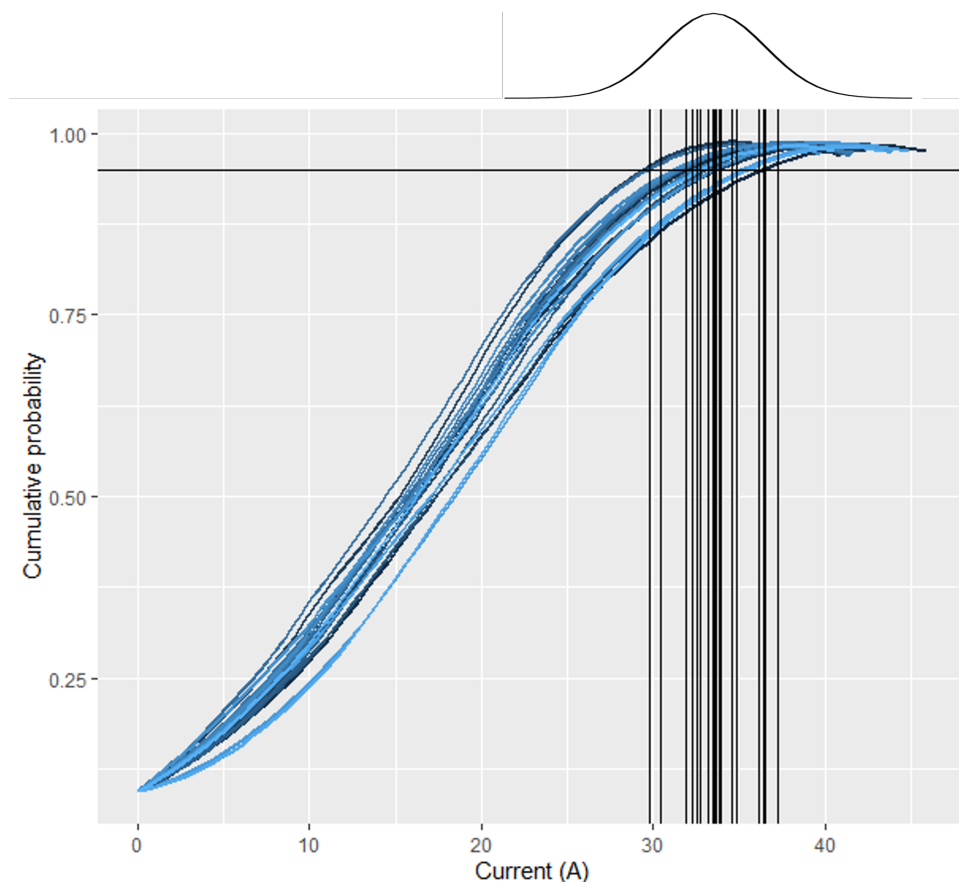
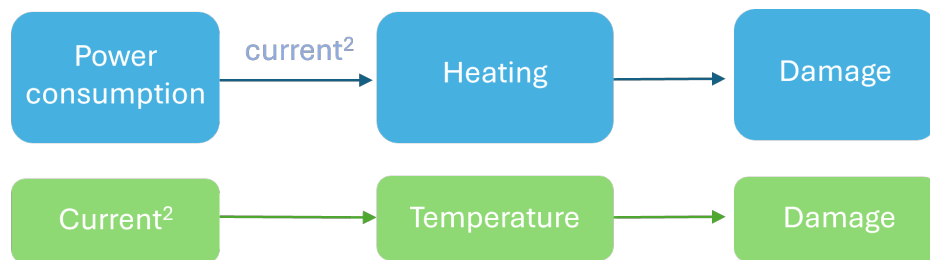


Figure 43: Cumulative distribution of current for many samples. Vertical black lines show the extracted q95 value for each sample: the normal distribution provides a good fit to the set of these values.

The physical process of cable and joint degradation from thermal stresses follows a clear pathway. Demands on the network from customer consumption cause currents in the cables; currents cause resistive heating both in the cables themselves and at the joints through effects such as contact resistance; and finally heating of cables and joints leads to accumulation of damage, both through prolonged high temperatures and temperature cycles.

The mechanisms and stages of the damage processes and the factors that influence current, temperature and damage individually are well understood so it is reasonable to mirror this structure in the statistical modelling chain, as we can use this knowledge to select appropriate model forms and features. First a statistical model that predicts current was trained with only basic and known network properties as input features. The predictions from this model

are then used as the inputs of a second model, predicting temperature from current. Finally, these temperature predictions are fed into a model to predict damage as depicted in Figure 44.



Physical degradation paths inform the structure of statistical models

Figure 44: Data flow of the statistical models

6.0.3 Out-of-sample model evaluation

It is important to evaluate the performance of the fitted statistical models, to assess their suitability and skill. Evaluating their predictive performance on points that were also included in the training data would result in artificially inflated performance, not reflecting real performance on unseen data. Out-of-sample evaluation is performed to avoid this, using cross validation⁴ to maximise use of the dataset and obtain out-of-sample predictions for all data points. The relationship between demand, current, temperature and damage is likely to be different for each individual cable and network. To avoid information leakage in the model evaluation, all simulations for a given cable are assigned to the same fold. The same fold labels are used for all models, so the fold label for a given cable also applies to the joint that it is connected to.

6.1 Current models

The first model in the chain of statistical models predicts (quantiles of) current (I) per cable. Looking ahead to the next model in the chain, we know that temperature is related to the *square* of current — not current itself. Therefore, direct predictions of quantiles of I^2 will be more informative for the subsequent modelling and likely lower error compared to predicting quantiles of I and then squaring the prediction.

As described above, we know current is normally distributed. Along with the fact current must be positive, we can say that the square of current is lognormally distributed:

$$\begin{aligned} I &\sim \mathcal{N}(\mu_I, \sigma_I) \\ \log(I^2) &\sim \mathcal{N}(\mu_{I^2}, \sigma_{I^2}) \end{aligned} \tag{28}$$

In LV Predict 1, a parametric model was used, where the parameters μ_{I^2}, σ_{I^2} were predicted — these two parameters completely describe the distribution for each current quantile. However, in LV Predict 2 we found this approach to be less stable than desired, for two reasons. Firstly, predictions for μ_{I^2} and σ_{I^2} are made on the log

⁴ see Section B.3 for full details

scale and then transformed (through exponentials) to give parameters for I^2 . The exponential transformation means a relatively small overprediction of a high value in log-space becomes exponentially larger when transformed: the actual transformation is slightly more complex but conceptually, a predicted value for the standard deviation of $\log I^2$ of 8 would result in a standard deviation in I^2 that is proportional to e^8 . Thus, relatively small errors in predicted values in 'log space' are magnified under the transformation, particularly for higher values, resulting in extreme and unrealistic predictions for the standard deviation of I^2 . Under this model, the parameters μ (a measure of the central estimate of I^2) and σ (a measure of the spread of uncertainty) are passed to the next models (temperature) in the statistical modelling chain. To avoid this issue of the scale of predicted values blowing up through the reverse log transformation, an alternative approach was implemented. Models were trained to predict quantiles of I^2 directly without the log transformation, with the central estimate provided by a deterministic model and the uncertainty estimate derived from the difference between two quantiles. The central 30% interval (i.e. between q35 and q65) was chosen as the measure of uncertainty.

Examining the relationships seen in the simulated training data can help to understand what features are most informative, and the expected form of those relationships. Figure 45 shows a relatively strong relationship between the number of customers downstream and quantiles of the resulting current in the cable. The colouring also shows a relationship between current and LCT penetration (proportion of EVs and heat pumps) for a given number of customers downstream, suggesting this is also informative. EV and heat pump features are included independently of each other in the model, as the different technologies have different demand profiles and therefore different relationships to current.

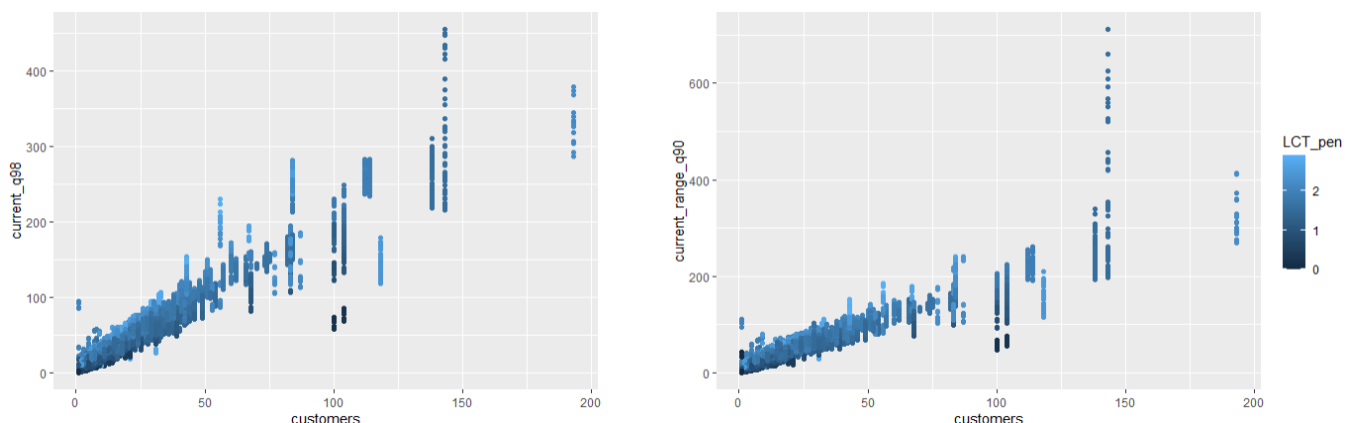


Figure 45: Relationship between number of customers downstream (x) and current quantiles.

The ProbCast package in R was used to fit this model, as it has inbuilt capability to model a central estimate plus quantiles of uncertainty through the `qreg_gam` function. This fits a GAM model for the central estimate, and then fits quantile models to the residuals of this central prediction. A GAM was chosen for the central model as it fits smoothly-varying, nonlinear relationships between the input features and the target which are appropriate for the relationships between network features and current⁵. The smoothness of these relationships and the interactions between input features can be controlled by specifying the form of the splines. The form for the central estimate is the same for each quantile $q = \{q95, q98, q90_{range}\}$:

⁵ see B.1 for more details

$$\log(I_q^2) = A + s(c) + s(E) + s(N) + s(P_{EV}) + s(P_{HP}) + ti(c, E) + ti(c, P_{EV}) + ti(E, P_{EV}) + ti(c, P_{HP}) + ti(E, P_{HP}) \quad (29)$$

where c is the number of customers downstream of the modelled cable, E is the total annual energy of the downstream customers, N is the total number of customers on the feeder (a measure of network complexity), P_{EV} is the EV penetration of downstream customers, and P_{HP} is the heat pump penetration of downstream customers. $s(\cdot)$ are one-dimensional smoothly varying relationships between one input feature and the target, while $ti(\cdot)$ are two-dimensional smooths that allow for interactions of two input features⁶. A is an intercept term. Using a log-link function — modelling the log of I_q^2 — ensures that all predicted values will be > 0 since I_q^2 is the exponential of the linear combination of features.

The models for q35 and q65, which together are used to derive a predicted uncertainty feature, are linear models with the form

$$q_{35}(I_q^2) = c + N + P_{EV} + P_{HP} \quad (30)$$

Figures 46 to 48 show the fitted marginal relationships between individual input features and the predicted 95th quantile of I^2 — that is, the effect that the input feature on the x axis has on the target variable when all other input variables are fixed. These plots show the shape of the fitted relationships for the $s(\cdot)$ terms in Equation 29. Note the difference in scale between the figures: the number of customers downstream of the target cable and their total annual energy have much larger magnitude contributions to the prediction than the network size (total feeder customers) or EV and heat pump penetrations. As expected, cable current increases with number of customers downstream. For lower values of annual energy (the total annual energy demand of all downstream customers), there is a positive relationship to cable current, as might be expected, up to values of around 200 MWh/year. At higher annual energy demands than this, the marginal contribution to the prediction from annual energy decreases. Very high annual demand likely relates to cables with a high number of customers and since larger groups of customers have more demand diversity this is likely to result in a lower current relative to the total annual demand. Additionally, high energy consumption could be the result of consistently high demand, not necessarily increasing the peak demands (and therefore not increasing the high quantiles of current). Network complexity as measured by the total number of customers on the feeder and also has a varying relationship with current, with generally higher currents for larger networks. Physical cable properties - mainly cable conductor area - will affect the cable's resistance and therefore the resulting relationship between demand and current. These physical cable properties also vary by both network size and point on the network. As might be expected, higher EV penetration is related to higher q95 I^2 , but the inverse relationship (decreasing q95 I^2 with increasing penetration) is seen for heat pumps. This may be due to the timing of these two demands: EVs tend to be plugged in in the evening, increasing peak demand and therefore increasing q95 I^2 , whereas heat pump demand is distributed more evenly across the day with peak heat pump demands tending to fall in the morning instead of coinciding with peak evening demand.

⁶ See appendix B.1 for more details on the form and construction of these spline terms

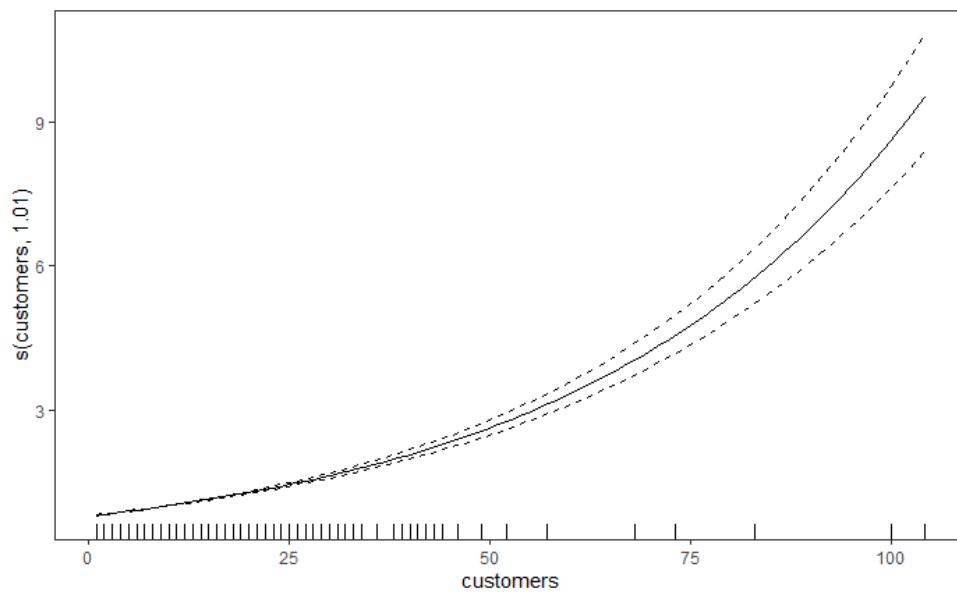


Figure 46: Fitted marginal relationship between number of downstream customers and $q95 I^2$.

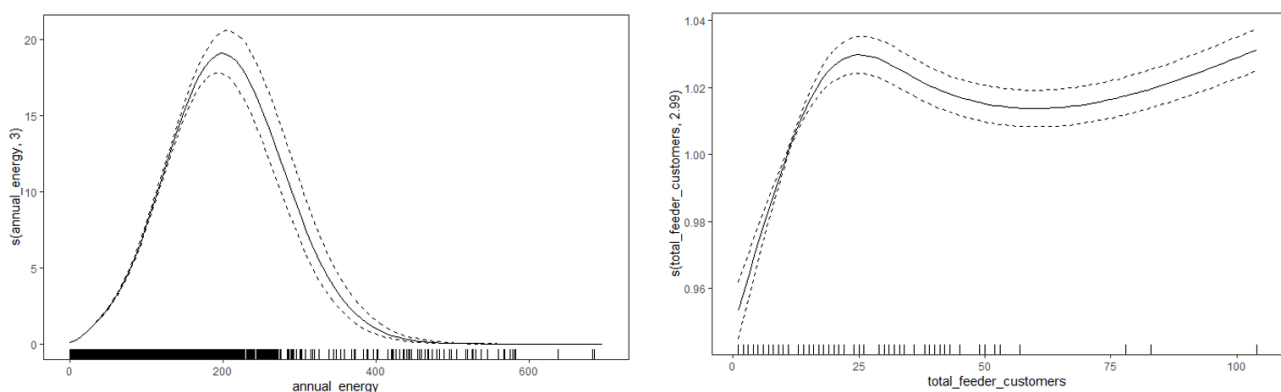


Figure 47: Fitted marginal relationship between per-customer annual energy consumption of all downstream customers and $q95 I^2$ (left), and total number of customers on the feeder - a measure of network size - and $q95 I^2$ (right).

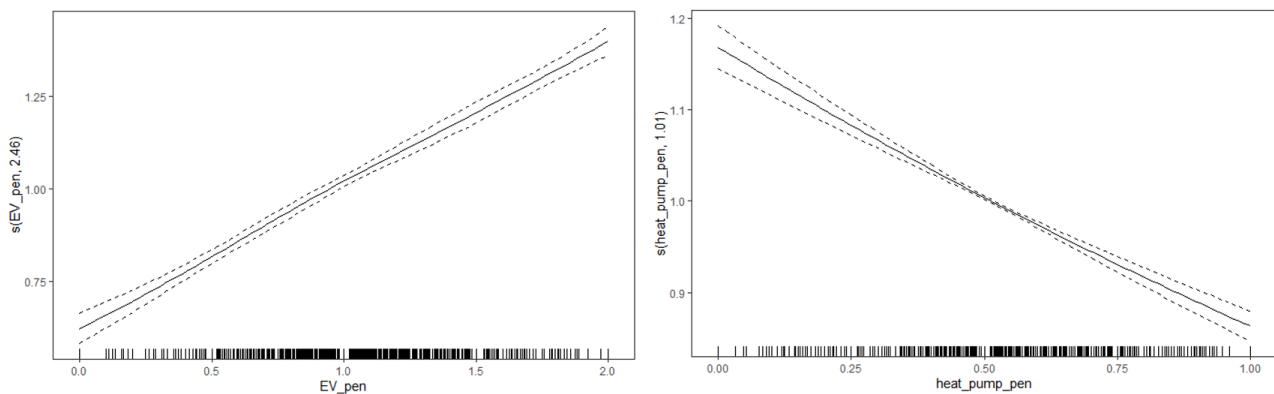


Figure 48: Fitted marginal relationship between EV penetration of all downstream customers and $q_{95} I^2$ (left), and the equivalent for heat pump penetration (right).

As well as the marginal relationships between individual input features, the fitted two-dimensional spline forms can be plotted. Figure 49 shows the marginal effect of the interaction term between number of customers and total annual energy. A large positive marginal effect (higher predicted value of $q_{95} I^2$) is seen for cables with a very large number of downstream customers and high total annual energy. Figure 50 shows the fitted interactions between downstream EV/heat pump penetrations and the number of downstream customers. There is little additional effect from the interaction term for most combinations of variables. For both EVs and heat pumps, the highest currents are predicted for the cables with the highest total number of LCTs (i.e. a high number of customers and EV penetration).

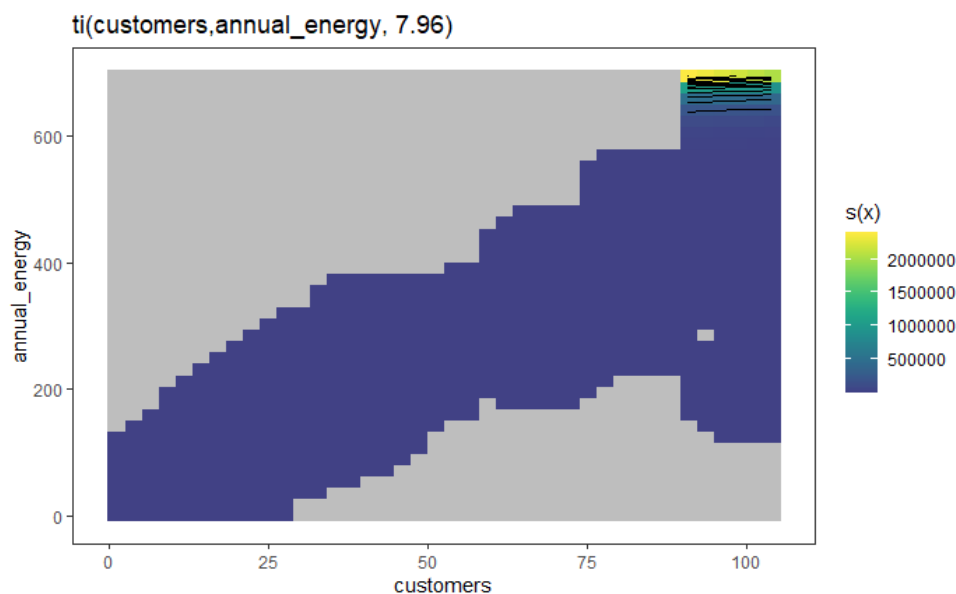


Figure 49: Fitted marginal interaction term between number of downstream customers and their per-customer annual demand. Colour shows the size of this term's contribution to the prediction of $q_{95} I^2$.

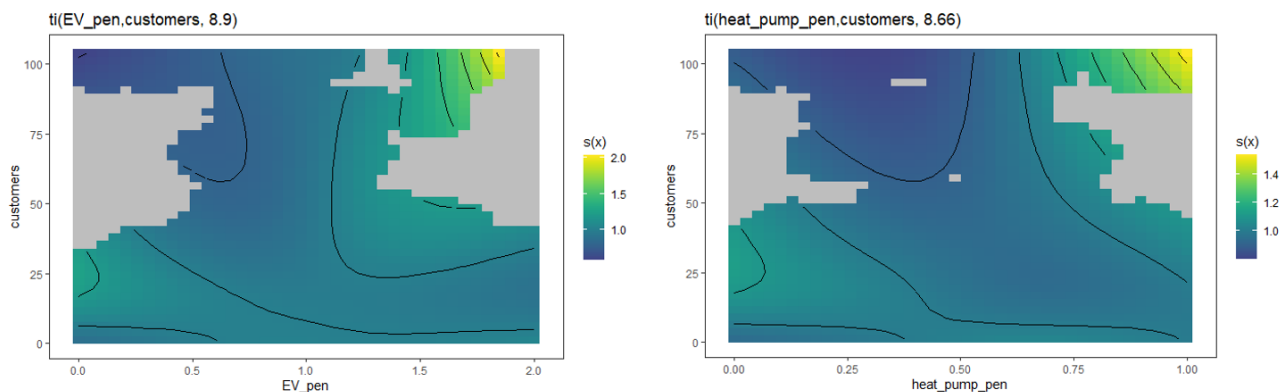


Figure 50: Fitted marginal interaction term between EV penetration (left)/heat pump penetration (right) and number of downstream customers. Colour shows the size of the term's contribution to the prediction of q95 I^2 .

6.1.1 Smart meter features

Intuitively, knowing the exact half-hourly demand of a subset of customers should provide additional information that gives an improvement in the predictions of current for cables on that network. In LV Predict 1, predictions of demand were made separately for the smart meter customer subset, compared with the actual values and this was used to 'correct' the overall prediction. This resulted in sharper (more confident) predictions of overall current. This adjustment followed this main step:

- Make predictions of current quantiles for the cable with all demands, and for the cable if only the smart meter demands were present.

In LV Predict 1, each cable was treated independently as a single section with all downstream customers treated as an aggregate load, and using a simple linear conversion from demand to current. This meant the known smart meter demands could be thought of as a proportion of the overall demand, and treated accordingly. The linear conversion from demand to current also meant that the relationship between smart meter demand and current could be treated in the same way as the relationship between total demand and total current, and predictions of the proportion of current caused by the smart meter demands could be predicted from the same model as total current. However, this assumption does not hold when loads are assigned to specific locations on a network and a load flow calculation run to determine the current resulting from them.

Furthermore, it was assumed that at the point of model prediction, the network topology and exact location of smart meter demands would not be known, and therefore, the only smart meter data recorded was the total demand from all smart meter customers on the feeder. This feature has a much weaker relationship to current in any given cable, as some of those smart meter demands may be upstream of the cable of interest.

The inclusion of several variations of features based on the feeder-level smart meter demands was tested, and a description of the model term and its effect on overall performance is described in Table 7. Here sm_{q95} is the 95th quantile of smart meter demand, used in the model for q95 current — the 98th quantile is used for q98 current, etc. P_{sm} is the smart meter penetration (the proportion of customers on the feeder who have smart meter data). As seen in the table, none gave an improvement in the predictive skill of current as compared to the model form in 29.

Included term	Resulting increase in error
$ti(sm_{q95}, P_{sm})$	22%
$ti(sm_{q95}, N)$	22%
$s(sm_{q95} * N / N_{sm})$	7.5%

Table 7: Tested smart meter terms in q95 current model and their impact on model error.

If the network topology and locations of all smart meters was known for all networks, it would be possible to record and include downstream-only smart meter demand as a feature in the current model, which may be more likely to provide additional skill to the model. In order to improve the sharpness of the predictions (but without compromising their reliability), smart meter features should be included in the model form for the uncertainty predictions (Equation 30).

6.1.2 Current model evaluation

The skill of both the central current predictions and the quantile predictions is assessed for all target variables: q95 and q98 current, and q90 of daily current range. The central estimate can be treated like a deterministic prediction, and metrics of MAE and Mean Absolute Percentage Error (MAPE) are used to assess its accuracy. An initial examination of model errors showed significantly poorer performance at a small subset of networks. Investigation showed these networks had one main feature in common: they all had three or more feeders, where one feeder had a much larger number of customers compared to the other feeders. For example, one of these networks has three feeders with 143, 44 and 27 customers. This may result in more complex power flows, where the simple summary features of customer demand are less strongly related to the extremes of current seen. An indicator feature for these 'complex' networks was created, where a network is defined as 'complex' if it has more than two feeders and the largest feeder has at least 50 customers more than the next largest feeder. A model where this indicator variable was included as a feature was tested, and this would allow a single shift in intercept in each of the models to allow for differences in current in the complex networks. This approach gave a 2.4% reduction in MAE for the 'normal' networks, but a 3.8% increase in MAE for the 'complex' networks. It is likely that the actual relationship between input features and current is different for the complex networks, so instead, separate models were fitted for the 'normal' and 'complex' networks. This gave a reduction in MAE of 64% for the 'normal' networks, and a reduction of 16% for the 'complex' networks. While much larger errors are still seen for the 'complex' networks as shown in Figure 51, this is a significant improvement. Errors shown are on the scale of current squared. Inclusion of features that describe the network topology and impedance could help improve the model performance particularly for the 'complex' networks; alternatively, given the number of these is relatively small (4 of 111), running full physical simulations for these networks instead of using statistical models could be considered.

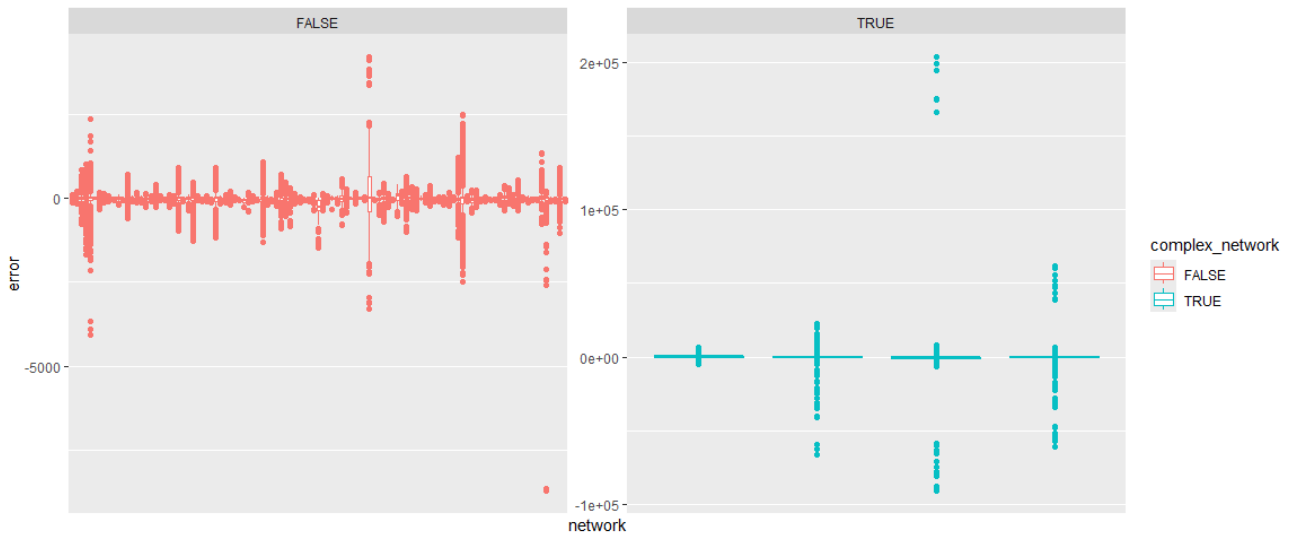


Figure 51: Box and whisker diagram of error in prediction of q95 I^2 per network. Note the difference in y-axis scale for complex network errors.

The relationships between predicted mean value, and simulated value of current quantiles, are plotted in Figures 52 to 54. In general, there is good agreement between the statistical model predictions and simulated current quantiles, particularly for the 'normal' networks.

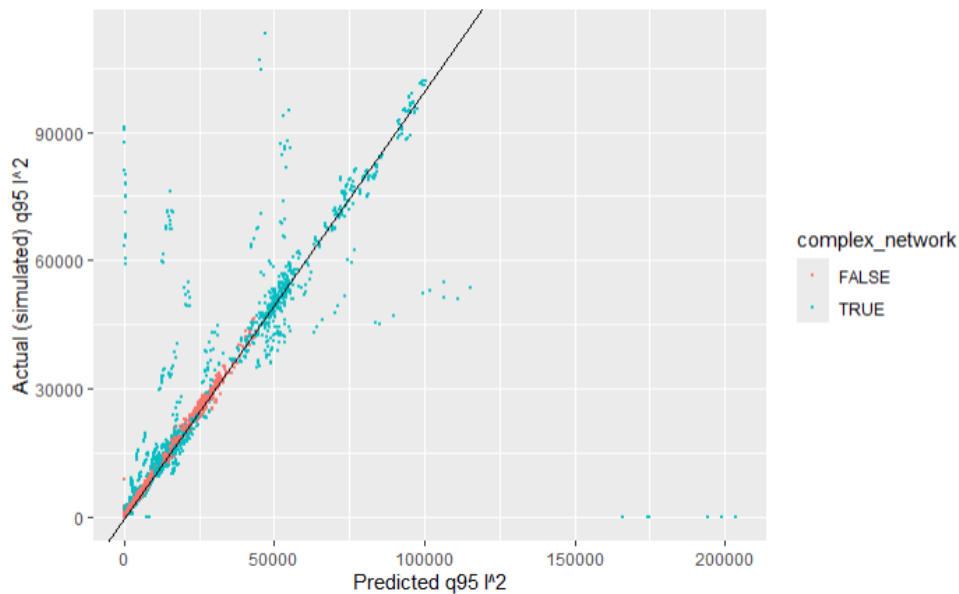


Figure 52: Relationship between predicted and actual q95 current values.

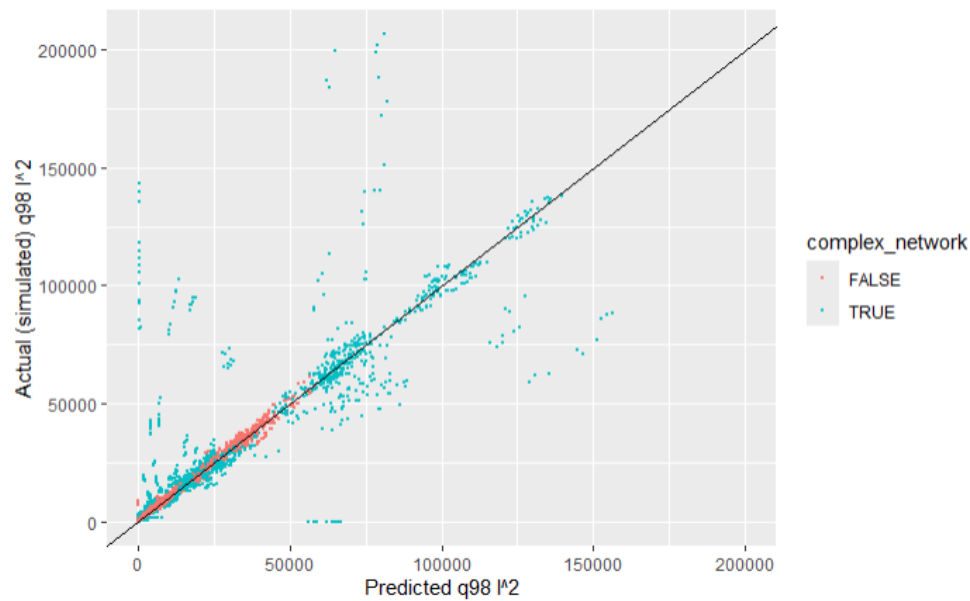


Figure 53: Relationship between predicted and actual q98 current values.

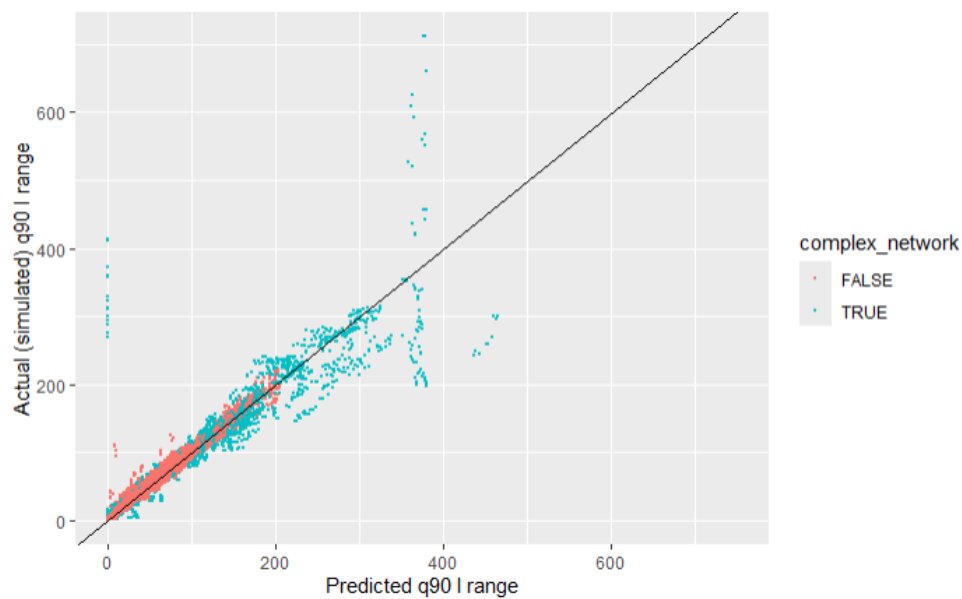


Figure 54: Relationship between predicted and actual q90 daily current range values.

These scatter plots only consider the skill in the central estimate of current quantiles, and not the uncertainty metric. The skill in the uncertainty metric is evaluated by checking the reliability of the model predictions. That is, the predicted q35 of, for example, q95 current, should be exceeded 65% of the time (and the q65 prediction should be exceeded 35% of the time). These expected vs observed exceedance frequencies are plotted in 55, and show good reliability and meaning the predicted uncertainty metric is calibrated.

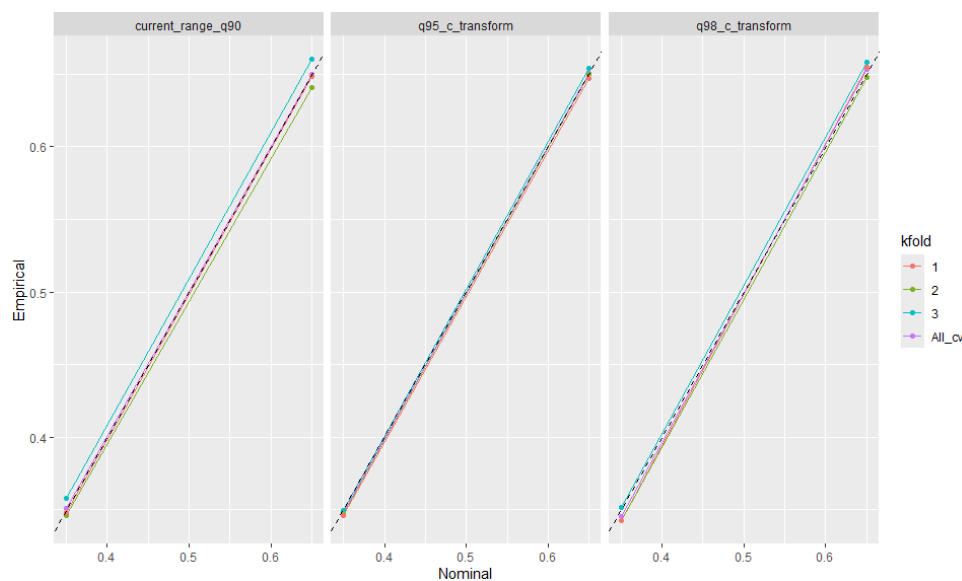


Figure 55: Reliability diagrams for the q35 and q65 predictions. Perfectly reliable predictions would lie on the black $y=x$ line.

6.2 Temperature models

Once predictions of current have been made, these are used as inputs to the next model in the chain — predicting quantiles of joint temperature. Quantiles of the *square* of current per cable were predicted, since temperature is proportional to I^2R/L . The predictions of I^2 were scaled by R/L , the cable resistance per unit length, to get a predicted quantity that is theoretically proportional to the heating effect of the current in that cable. Since multiple cables are connected at one joint, there are multiple predicted current values around the joint. The physical temperature model simulates temperature at several locations within the joint, and the returned peak temperature is the maximum across the locations modelled. Similarly, the largest predicted current from all of the connected cables was selected to provide the input features for the statistical models for temperature. Properties of the joint are also summarised into features for the number of cables around the joint, and the proportion of connected cables that are XLPE (as opposed to PILC).

First the relationships between current and temperature are explored. This is based on the output currents and temperatures of the physical models to clearly see the relationships present, without the additional variability from imperfect predictions from the statistical model for current obscuring the relationships.

Figure 56 shows the relationship between 95th quantiles of transformed current (I^2R/L) and temperature. The effect of higher elevation to lower temperatures for the same current quantile can be seen in the colour of the points. This is due to the effect of higher elevation lowering soil temperature. Surrounding soil temperature is also affected by latitude, the soil thermal resistivity which is a property of the soil type and local rainfall conditions, and recent air temperatures. While air temperatures are not included as features in this model, a random effect for the year of each data point is included. This encapsulates year-to-year differences in soil temperature. Figure 57 shows the relationship between corresponding current and temperature daily range features. Here the current feature is not squared, as the value is already the difference of two values and $(a - b)^2 \neq a^2 - b^2$. For the same reason, this cannot purely be scaled by the factor R/L to create a current feature that is linearly proportional. However, this plot

shows the quadratic relationship between current range and temperature range, and how the factor R/L shifts this relationship.

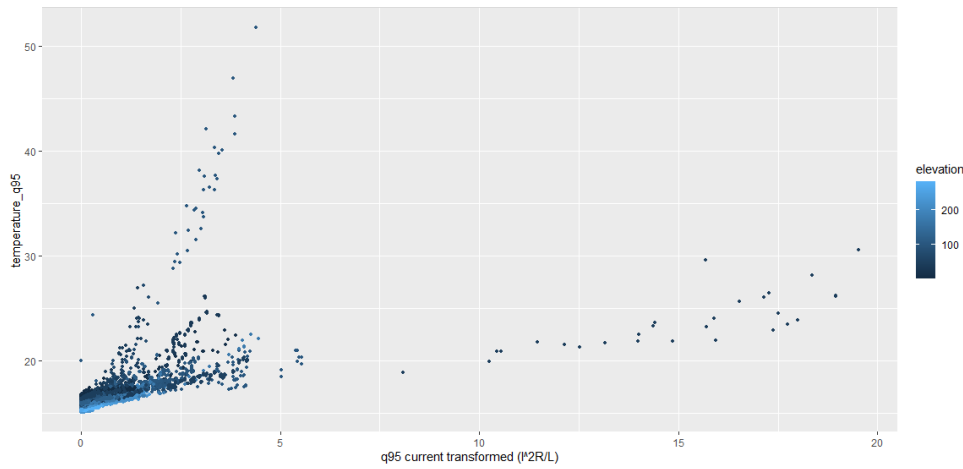


Figure 56: Relationship between transformed current feature and q95 temperature.

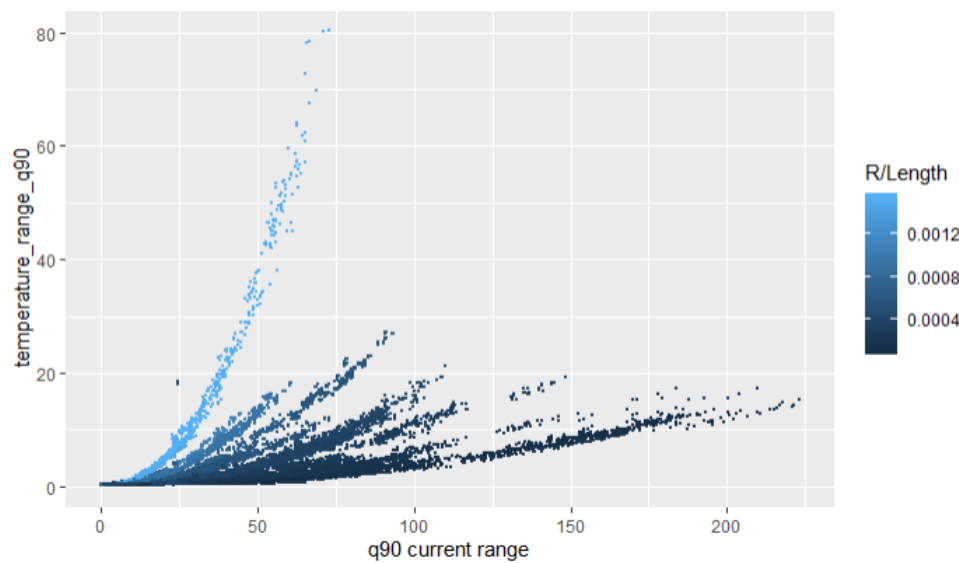


Figure 57: Relationship between transformed current range feature and q90 temperature range.

q95, q98 and q90 range of temperature are all normally distributed, and so are modelled as a Gaussian with mean and standard deviation parameters using the call `gam(family='gauss')` in the R package `mgcv`. We specify separate model forms for the mean parameter μ and standard deviation σ . For the models for q95 temperature and q98 temperature, the model forms are:

$$\mu_{q95} = s(q95(I^2R/L)_\mu) + s(e) + s(l) + s(\rho_{soil}) + re(y) + R + C \quad (31)$$

$$\sigma_{q95} = s(q95(I^2R/L)_\sigma) + C \quad (32)$$

where $q95(I^2R/L)_\mu$ is the predicted mean value of $q95 I^2R/L$, $q95(I^2R/L)_\sigma$ is the predicted uncertainty measure for $q95 I^2R/L$, e is elevation, l is latitude, ρ_{soil} is the soil thermal resistivity, y is the year, R is a rural/urban indicator variable and C is the indicator variable for normal vs complex networks. $s(\cdot)$ are one-dimensional smoothly varying relationships between one input feature and the target⁷, and $re(\cdot)$ is a random effect term - effectively allowing a separate intercept per year.

The form for the model for $q90$ temperature range is slightly different, as the predicted current input is not scaled by R/L directly so this has to be taken into account separately in the model. Because of this, the term $s(q95(I^2R/L)_\mu)$ in Equation 31 is modified to $ti(q90r(I)_\mu, R/L)$, i.e. a bivariate relationship between predicted current range and the scaling factor R/L . Similarly, the term $s(q95(I^2R/L)_\sigma)$ in Equation 32 becomes $ti(q90r(I)_\sigma, R/L)$.

The marginal fitted relationships for the terms in Equation 31 are shown in Figures 58 to 60. As expected, there is a positive relationship between the central estimate of I^2R/L and the predicted mean parameter for temperature. From a comparison of the scales of the y -axes in these figures, this term gives the greatest contribution to the temperature prediction as might be expected. Both elevation and latitude have negative relationships to temperature as is expected through their effect on soil temperature (i.e. higher elevation and higher latitude locations are generally colder). The relationship between soil thermal resistivity and temperature is less straightforward. The high uncertainty on most of the range of x values is due to the concentration of most soil resistivities at low values (indicated by the concentration of small vertical lines on the x -axis). The year-to-year variability that is not explained by other features in the model is contained in the random effects feature for year. There is slight variation year-to-year; limited conclusions on the long-term year-to-year variability can be drawn from the fitted random effect, given the data only spans three different years.

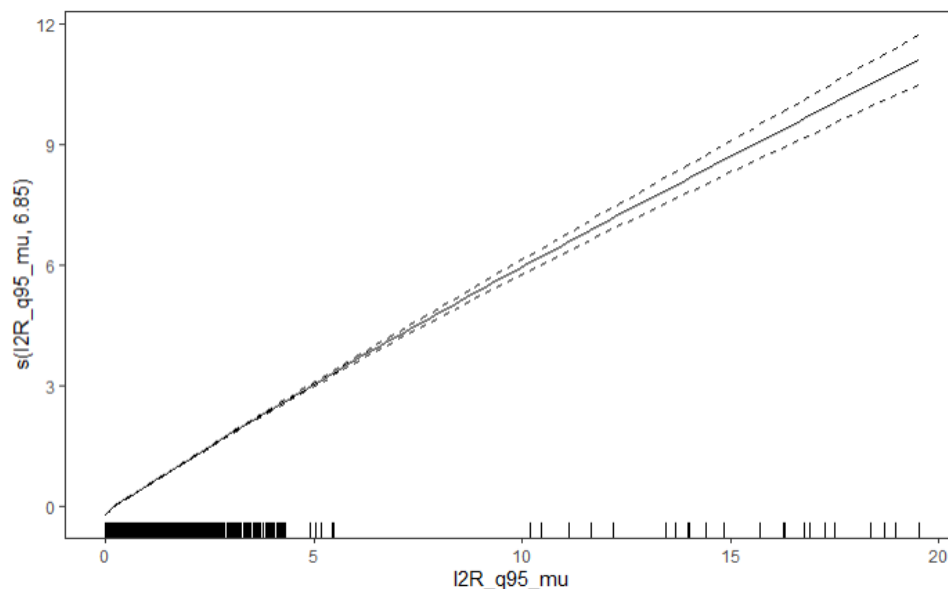
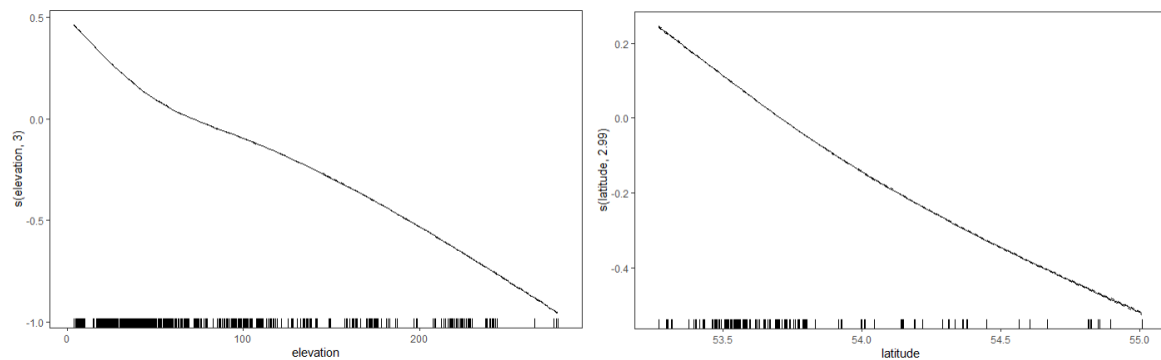


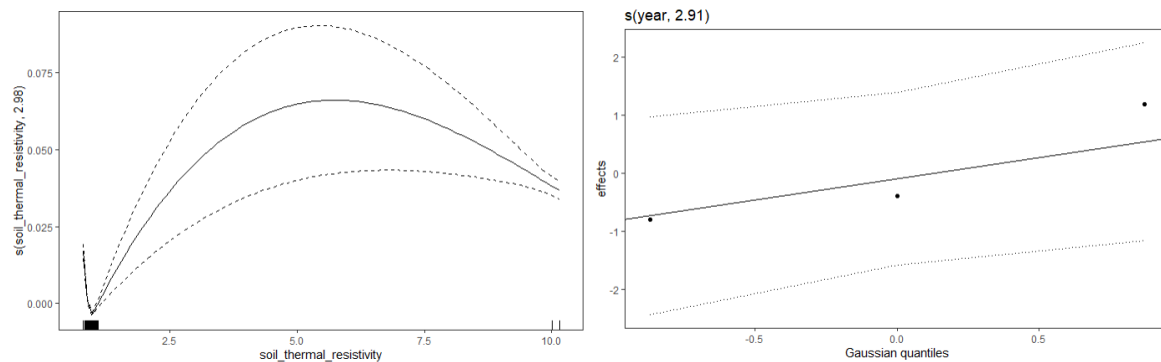
Figure 58: Relationship between the central estimate of transformed current, and $q95$ temperature.

⁷ See appendix B.1 for more details on the form and construction of these spline terms



a: Fitted marginal relationship between elevation and q95 temperature b: Fitted marginal relationship between latitude and q95 temperature

Figure 59: Fitted marginal relationships



a: Fitted marginal relationship between soil thermal resistivity and q95 temperature b: Fitted random effects for year in q95 temperature model

Figure 60: Further fitted marginal relationships

The relationship between predicted uncertainty in I^2R/L and the standard deviation parameter for temperature is generally positive, with most of the x values being below 0.125 (where the relationship is positive). A few samples have much higher uncertainty in I^2R/L which also translates into a higher value for σ for temperature.

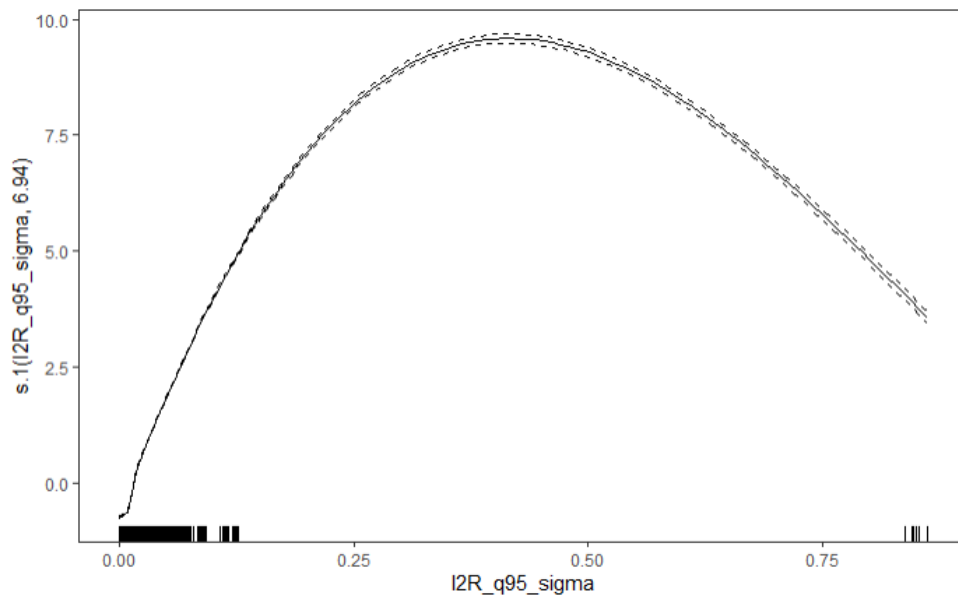


Figure 61: Relationship between q95 transformed current range uncertainty feature and the predicted σ parameter for q95 temperature.

6.2.1 Temperature model evaluation

The temperature models are evaluated in the same way as the current models. First looking at the predicted mean values for q95 temperature, again the 'complex' networks show much larger residuals (and much larger variability in these) as seen in Figure 62.

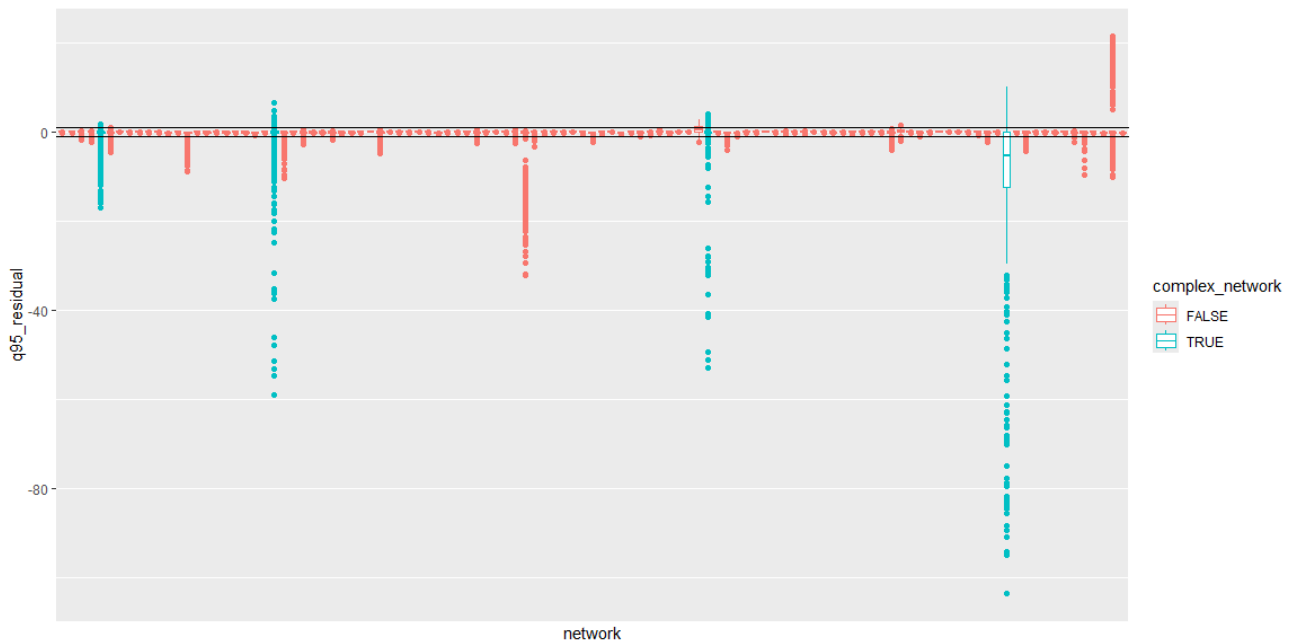


Figure 62: Box and whisker diagrams of error in prediction of temperature q95 per network. Black horizontal lines indicate a 1°C error.

The relationships between predicted mean values and simulated value of temperature quantiles are plotted in Figures 63 to 65. As seen in the predictions for current, the skill in the central prediction is much lower for the 'complex' networks. There are also some joints on the 'normal' networks where q95 and q98 temperature is under-predicted; however, 96.7% and 95.1% of samples respectively have a residual below 1°C. There are several samples where temperature range predictions are 0 when the actual q90 daily temperature range is not zero and vice versa. Further work may be needed to identify the cause of this. 87.8% of q90 temperature range predictions have a residual below 1°C.

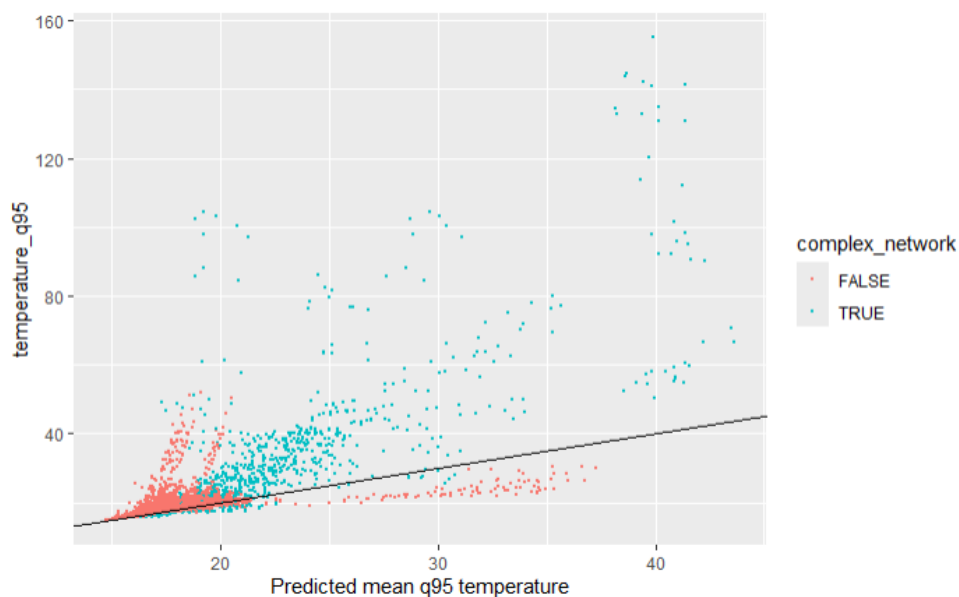


Figure 63: Relationship between predicted and actual q95 temperature values.

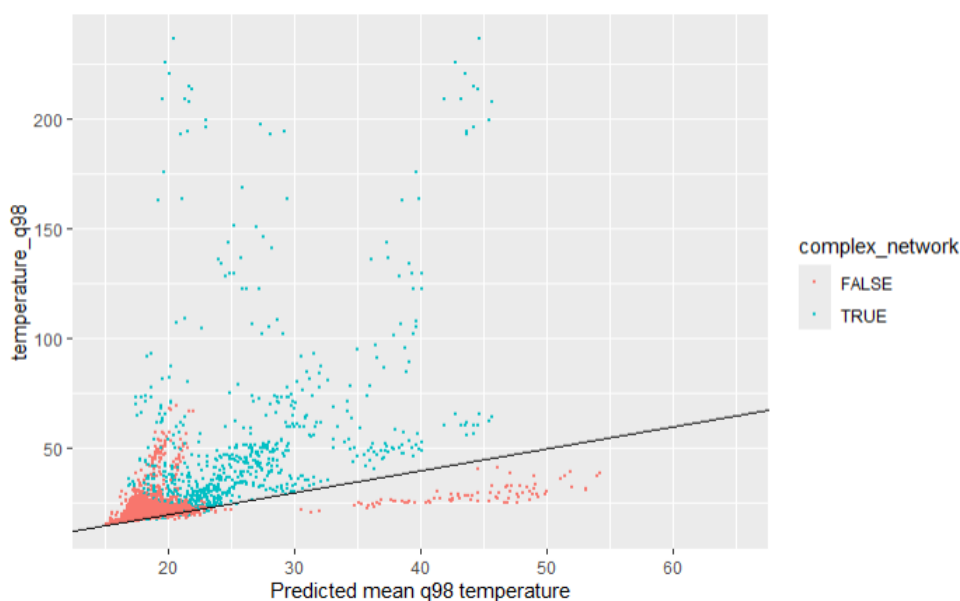


Figure 64: Relationship between predicted and actual q98 temperature values.

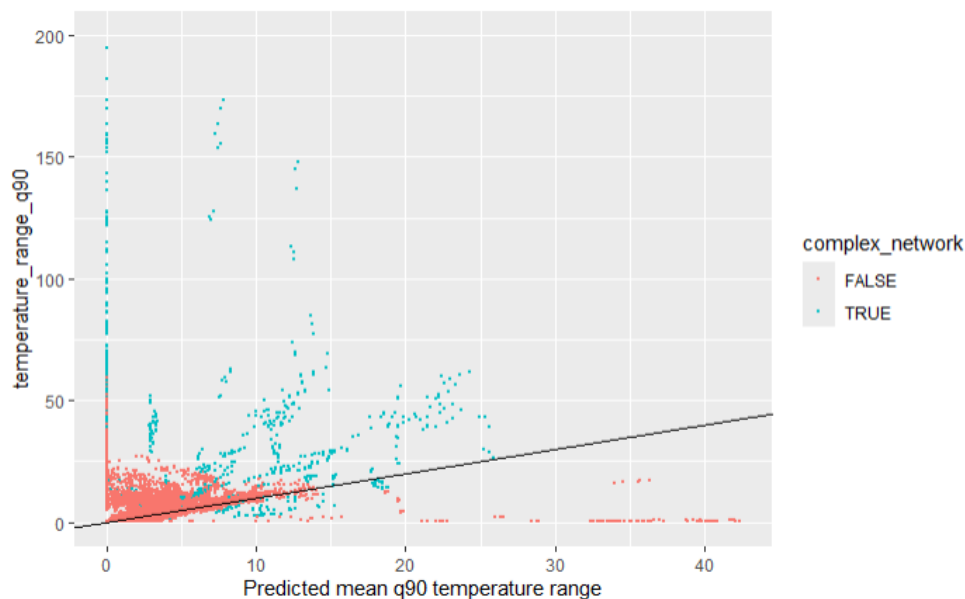


Figure 65: Relationship between predicted and actual q90 temperature range values.

The calibration of the whole probabilistic prediction of temperature quantiles, including both mean and standard deviation, was assessed through reliability diagrams as shown in Figure 66. Although not perfect, the reliability of the 'normal' networks is much better than that of the 'complex' networks. Both network types display an s-shaped curve, indicating the predicted distributions are too broad (and more so for the 'complex' networks), which is likely related to the much lower skill in predictions of current for these networks. Further refining of the model forms for temperature, in particular the form for σ , could improve this. Cable temperatures depend on several additional features that are not included as inputs in the current model form. Specifically, no metrics on ground temperature are explicitly included, only a per-year random effect to account for year-to-year variation in temperatures. Including features derived from actual ground temperatures, as well as features representing the level of sustained currents seen over time, could potentially improve the model reliability.

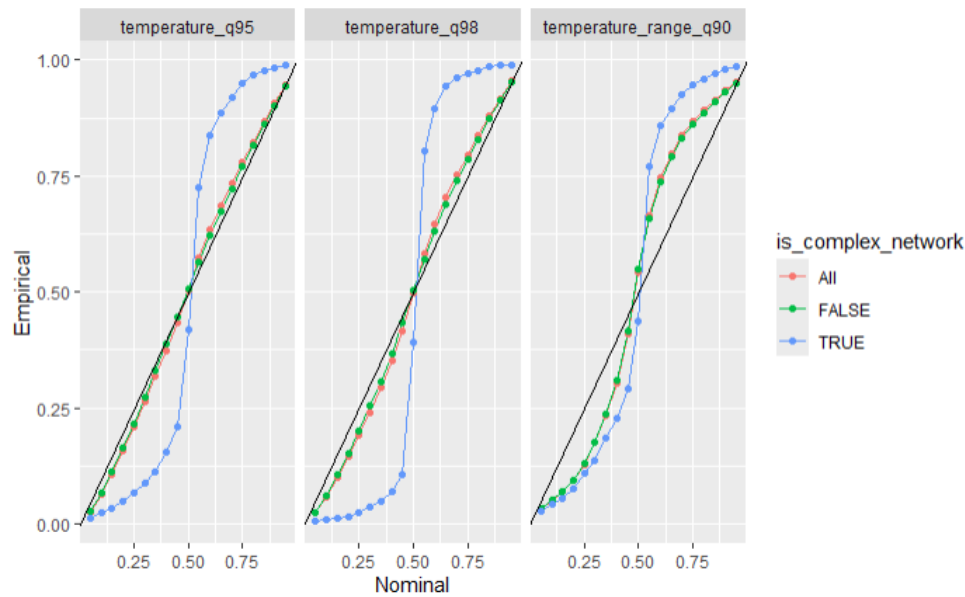


Figure 66: Reliability of probabilistic predictions of the three temperature quantile metrics.

6.3 Damage model

The final model in the chain of statistical models is damage. There are clear relationships between simulated temperature quantiles and damage, with different regimes depending on the mix of cable types around the joint as seen in Figures 67 to 69. In particular, very low damage is sustained for joints where all cables are PILC, as a bituminous joint filler is assumed instead of polymer. Creep, plastic and fatigue damage are only applied and accumulated in the polymer joints; the only contribution to damage in the bituminous joint is from chemical damage meaning lower damage for the same temperature conditions is expected in this case.

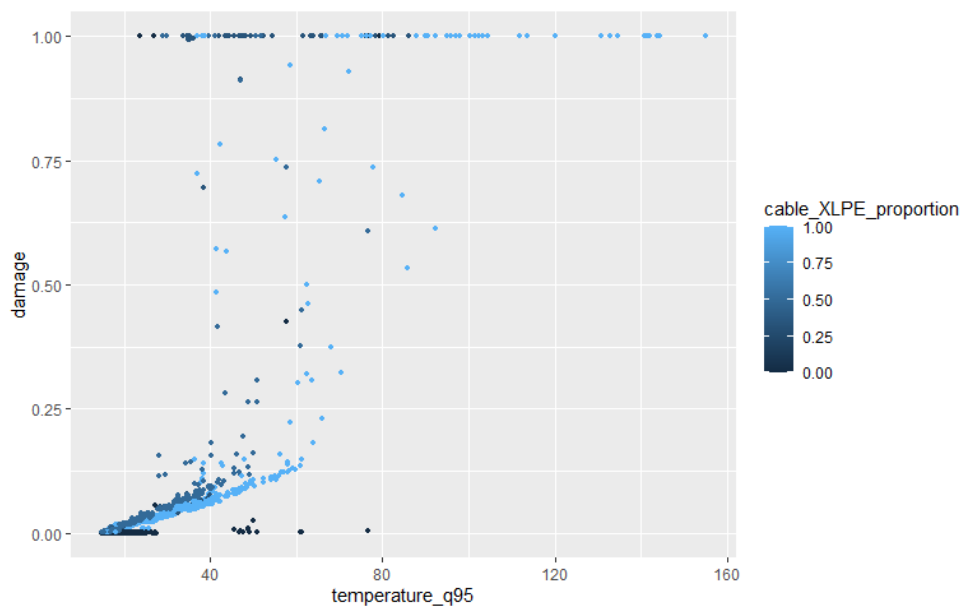


Figure 67: Relationship between simulated q95 temperature and damage.

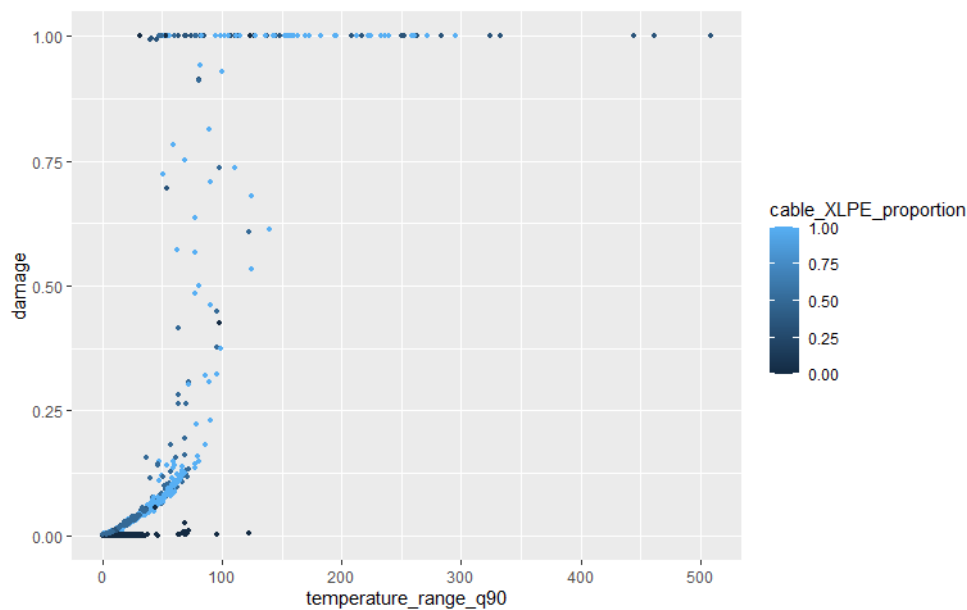


Figure 68: Relationship between simulated q98 temperature and damage.

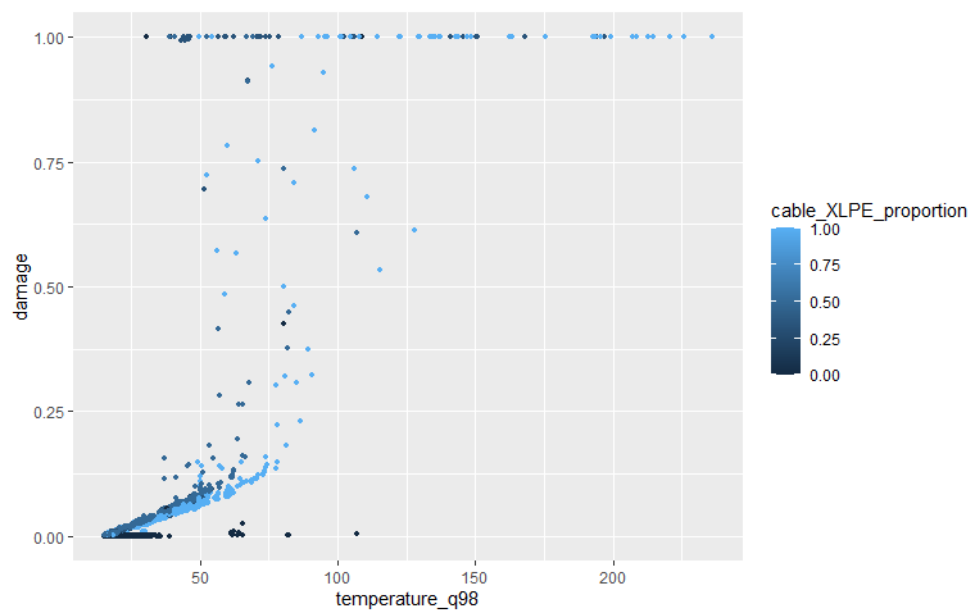


Figure 69: Relationship between simulated q90 temperature range and damage.

6.3.1 Initial model: beta distribution

Damage is only defined for the interval $0 \leq x \leq 1$, and the choice of predictive distribution must reflect this. The beta distribution was chosen as this is defined for the interval $0 < x < 1$ and provides flexible shaped distributions depending on the values of its two parameters μ and σ . To also allow for damage values of exactly 0 and 1, this distribution was modified to the *inflated* beta distribution, where additional probability masses are assigned at 0 and 1. The proportions of the total probability mass at 0 and 1 is controlled by adding two additional parameters to the specification of the distribution, ν and τ . The probability density for an example 0 and 1-inflated beta distribution is

shown in Figure 70.

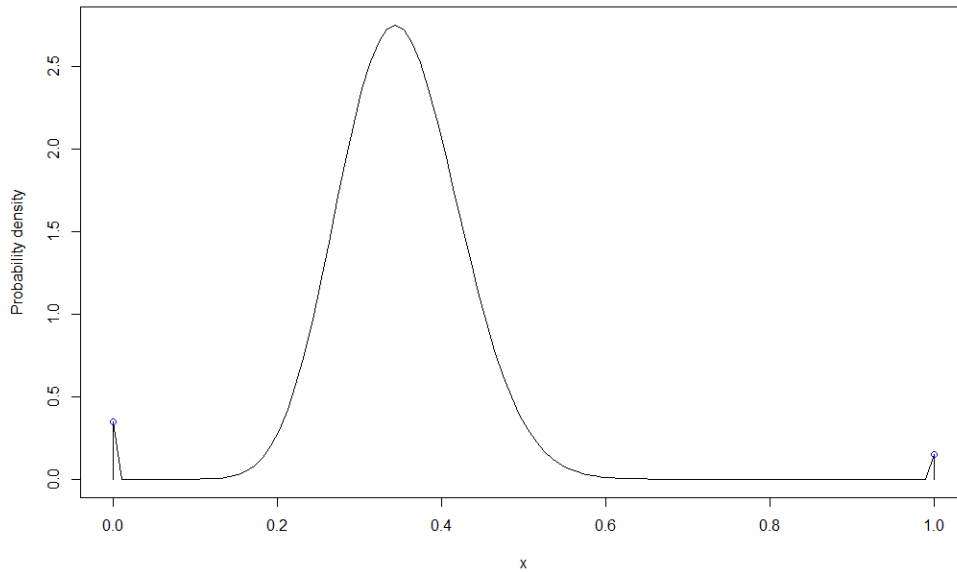


Figure 70: A 1-inflated beta distribution with probability mass on the boundary at 1.

The 0 and 1-inflated beta distribution is completely defined by 4 parameters: the mean and scale of the beta distribution, the proportion of probability mass at 0 and the proportion of the probability mass at 1. Model forms for these three parameters were specified as

$$\mu = t_{q98} + t_{q90range} + f_{XLPE} + N \quad (33)$$

$$\sigma = \sigma(t_{q98}) \quad (34)$$

$$\nu = t_{q98} \quad (35)$$

$$\tau = t_{q98} \quad (36)$$

where t_{q98} is the predicted mean q98 temperature value, $t_{q90range}$ is the predicted mean of q90 daily temperature range, f_{XLPE} is the proportion of cables connected to the joint that are XLPE. This value would be 1 for joints with only XLPE cables connected, and 0 for joints with only PILC cables connected. N is the total number of customers connected to the feeder, a measure of network complexity.

Fitting this model using the `gam1ss` package in R resulted in predictions with poor calibration as seen in Figure 71. The shape of these reliability diagrams does not show the distinctive shapes of either consistent bias (always under or over the $y=x$ line) or over-or under-confident predictions (an s-shape). Therefore, it is possible that the beta distribution shape is not appropriate for this application and so a non parametric modelling approach was also trialled for comparison. In addition, the `gam1ss` package's prediction functions can be quite unstable and often fail to return predicted values.

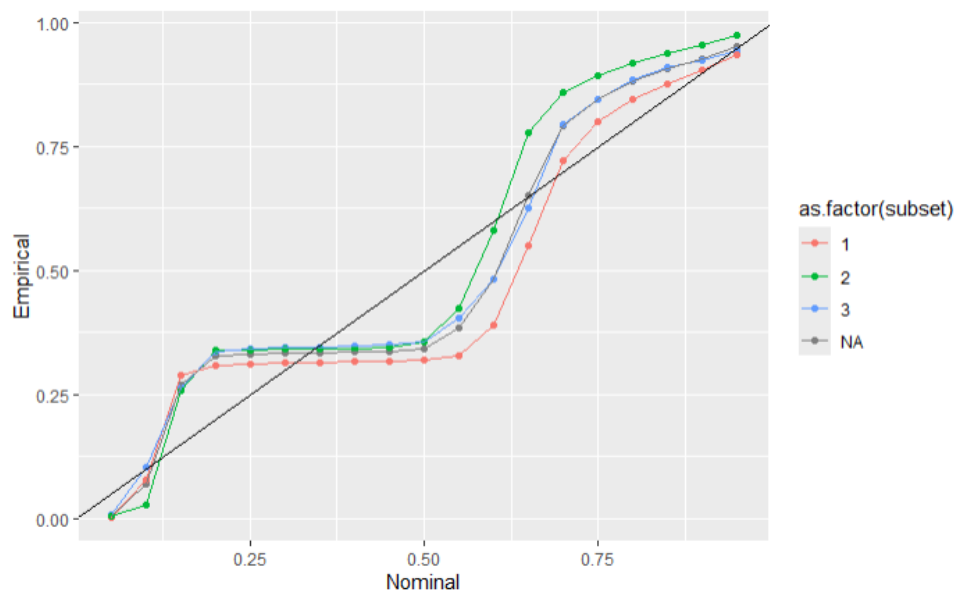


Figure 71: Reliability of damage predictions based on a 1-inflated beta distribution.

6.3.2 Refined model approach

Instead, the same approach as was used for the current model that was tested. This involves fitting a GAM for the central estimate of damage, and then fitting quantile regression models to the residuals of this central estimate to produce predictions for a set of quantiles that together define the predicted distribution. A very similar model form to that of the beta-distribution approach was used. The form for the central estimate was specified as:

$$\text{damage} = s(t_{q98}) + s(t_{q90range}) + s(f_{XLPE}) + N \quad (37)$$

Then a model for the residuals of these predictions was fitted with the single feature $\sigma(t_{q98})$.

As seen in Figures 63 to 65, while for most samples the predicted mean for temperature is very close to the simulated value, there are also cases with large residuals between predicted mean and actual temperature quantiles. To avoid the additional 'noise' from this degrading the fitted parameters in the damage model, the damage model is fitted on the *simulated* temperature quantile values — rather than their predicted means. This allows the damage model to fit to the 'true' relationship between temperature and damage without the additional noise from the imperfect temperature predictions obscuring these relationships. Then, when predictions are generated, the predicted mean value for temperature is used in place of the simulated value.

6.3.3 Damage model evaluation

All model evaluation in this section is based on the GAM plus quantiles model described in the previous section. The overall reliability of the damage predictions is extremely good, with the expected proportions of exceedances seen across the whole dataset as seen in Figure 72. However, since these predictions are based on the outputs of a model with less good reliability, it is worth investigating the actual proficiency of these predictions further.

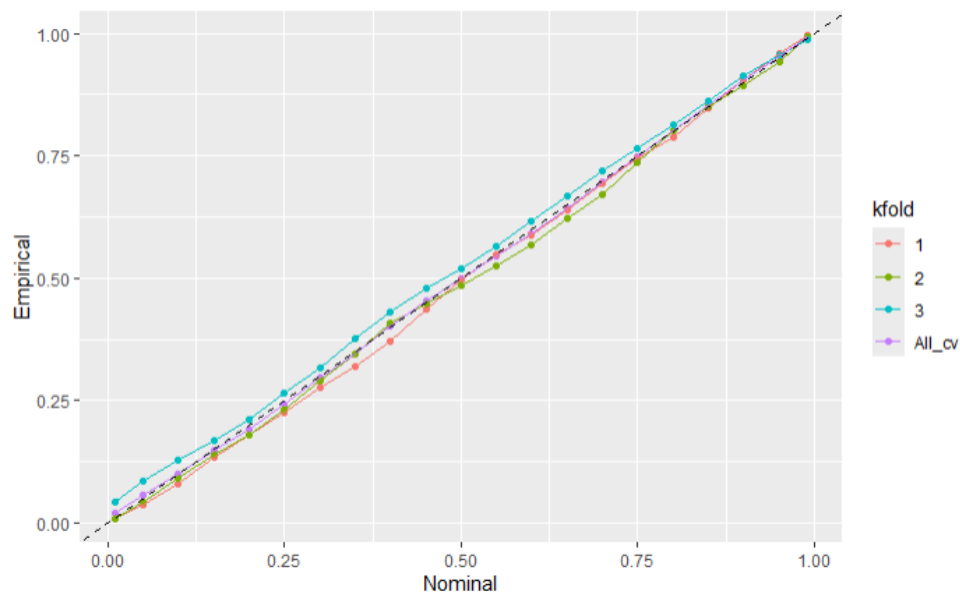


Figure 72: Reliability of damage predictions based on a GAM + quantile regression on the residuals.

The sharpness diagram in Figure 73 shows extremely sharp predictions, with the central 50% probability interval spanning a damage range of 0.0025 on average.

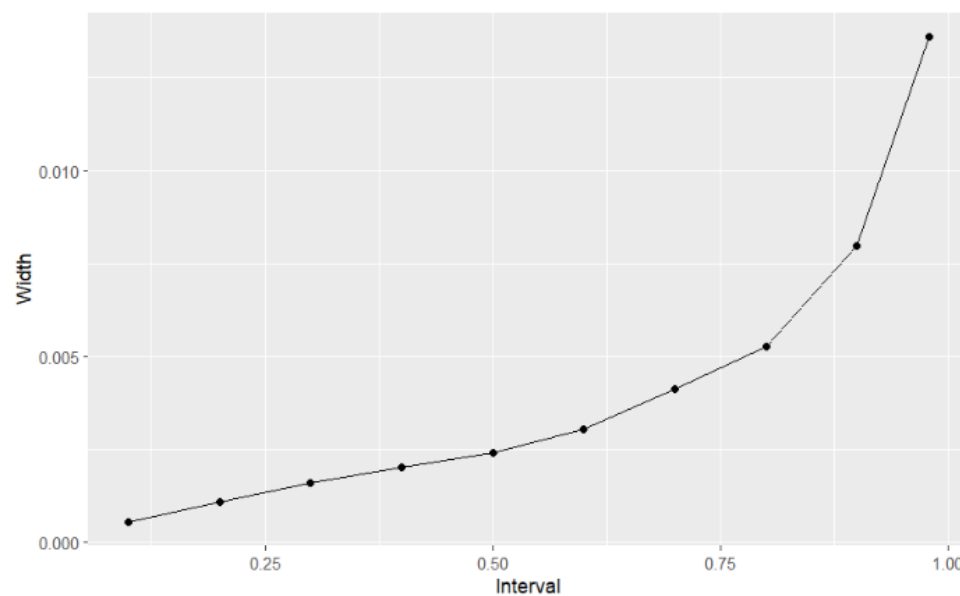


Figure 73: Sharpness diagram of damage predictions. x -axis is the predictive interval and y -axis is the width of this interval in terms of a range of damage values.

However, an examination of the median prediction of damage shows quite a large amount of scatter, and a high concentration of points with very low damage (< 0.025). Therefore, it appears the model is predicting more of an average distribution of damage, and is not able to effectively discriminate between cases of high and low damage in its predictions.

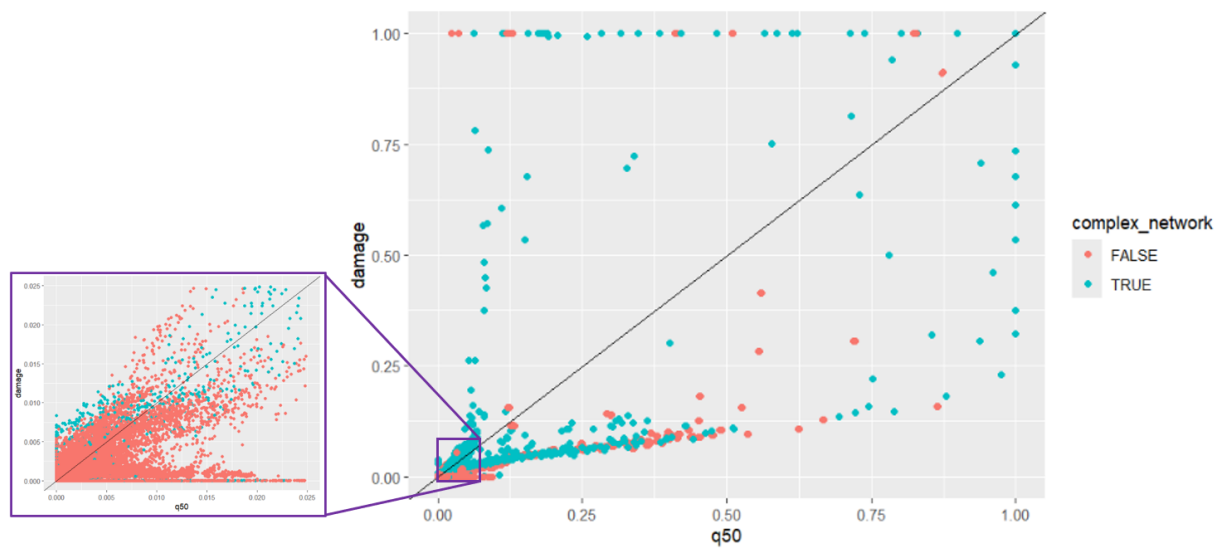


Figure 74: Relationship between predicted median damage and simulated (actual) damage values.

Assessing reliability by subsets of predicted median damage (Figure 75) shows that when the model is predicting very low damage (red and yellow lines), it is consistently under predicting (i.e. the true damage is often higher than predicted), and conversely for the highest 20% of predicted median damage values (the purple line), it is over predicting and damage is often lower than predicted. Since the overall model is reliable, this suggests that the supplied features are not informative in distinguishing high and low damage cases. Features expressing information about how temperature varies within the year (e.g. a feature for sustained high temperature as well as just quantiles of the half-hourly temperature) may help improve this. In LV Predict 1, good reliability was found for the damage model predictions; the change in skill of the final damage model is likely due to the additional complexity of the cable temperature modelling, and the increase in types of cables and joints considered meaning there is more variation in the relationships between currents, temperatures and damage across the population of cables. It is apparent that, due to the added complexity in LV Predict 2, there are no easily formulated, simplified representations of the complex physical processes that lead to cable degradation and damage.

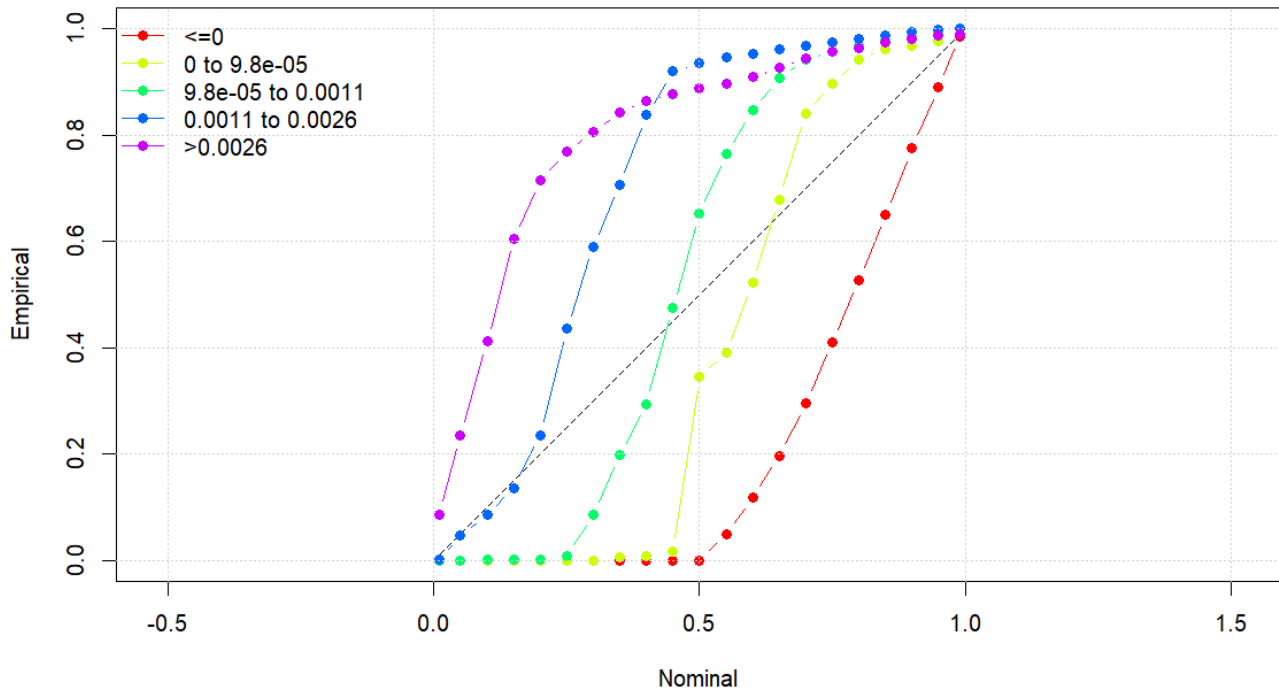


Figure 75: Reliability of predicted damage distributions, subsetting by predicted q50 damage value.

7 Model usage

The preceding chapters explain the steps of the methodology (sampling demand; defining expressions for the physical link from demand to temperature to damage; generating training data using these physically-based expressions; then fitting statistical models for current, temperature, and damage).

This chapter demonstrates how the outputs of this implemented methodology (fitted statistical models) can be used to make predictions for new ENW networks. The fitted models take the form of R objects that can be saved out when they are created, and loaded in for use later. Using fitted models for current, temperature and damage, it is possible to make predictions for new ENW networks based on only the inputs to the statistical models. This chapter demonstrates this process in detail for one example network, and then shows the results of applying this to a large set of networks.

While these fitted models reduce the amount of input data needed to make predictions, they remain complex and require familiarity with the associated model packages and mathematics for consistent use. A discussion of the practical considerations for implementing the methodology is presented in Section 9.3.

7.1 Single network example

The network from transformer 20624544 is used for this demonstration, shown in Figure 76. The results are illustrative, based on the innovation development model with known shortfalls (summarised in Sections 3.5 and 4.2) and have not been through a quality assurance procedure appropriate for making any business decisions from the results.

The GIS data is converted to a graph of network connectivity and joints where at least two separate LV — not service — cables meet, with customers connected, are selected as those to make predictions for. The number of customers at each service point is taken from ENW's dataset of customers per service point, and accumulated up the graph to get a total number of customers downstream for each cable. This is shown in Figure 77. Note that additional joints may exist on the network due to undocumented repairs, not accounted for in the available GIS data.

This network is in the Oldham local authority area, so EV and heat pump values are taken from ENW's Distribution Future Energy Scenarios (DFES) workbook [53] for 2030 combined with a number of households by local authority. This gave an EV penetration of 0.46 and a heat pump penetration of 0.077 for this scenario and area. The exact number of EVs and heat pumps at every individual service point was sampled from a binomial distribution where the probability is set to the EV or heat pump penetration and n is the number of customers at that service point. Average power at the service point was then adjusted upwards to account for the inclusion of these LCTs by sampling an average power for each EV and heat pump from the datasets for these technologies used in the sampling process. The total number of customers on each feeder was also calculated and recorded. This results in a per-cable dataset of the network shown in Table 8.



Figure 76: GIS data for network 20624544

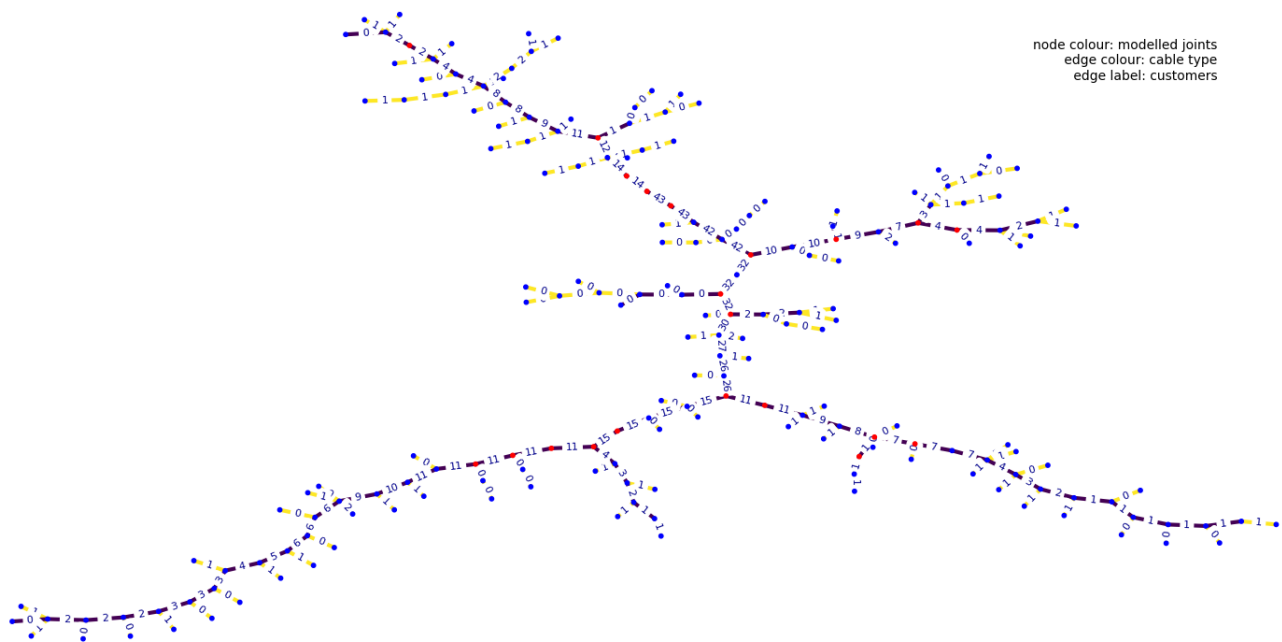


Figure 77: Network graph with accumulated customer numbers per cable. Red notes denote joints where predictions will be made. Yellow lines are service cables and dark purple lines are LV cables.

cable_fid	customers	EVs	heat_pumps	annual_energy	total_feeder_customers
8305293	43	19	3	173.53	43
7084215	14	8	0	73.41	14
7265609_1	14	8	0	73.41	14
7265609_0	12	8	0	66.55	14
7206532_0_0	11	7	0	61.81	14
9917827_0	1	1	0	4.73	14
7206532_1_1	2	1	0	3.96	14
9915947_0	2	1	0	3.96	14
7265607_0	43	19	3	173.53	43
7265607_1_1	42	19	3	168.19	43
7268450_0	32	14	2	137.32	43
7268454_0	10	5	1	30.86	43
7268454_1_0	10	5	1	30.86	43
5808114_1	9	5	1	29.66	43

Table 8: Table of network demand properties. Annual energy is in MWh.

For the current model, the required inputs per cable are:

- ▶ Number of customers downstream,
- ▶ Annual energy of downstream customers,
- ▶ Total number of customers on the feeder,
- ▶ EV penetration of downstream customers,
- ▶ Heat pump penetration of downstream customers,

- Complex network indicator (whether the network has three or more feeders where the largest feeder has at least 50 customers more than the next largest).

The complex network indicator is False for this network, which means the statistical model provides a reliable estimate of the current flows. Current is predicted per cable by loading the previously fitted models and supplying the features for the new network. Soil properties are determined from the cable locations and shapefiles of soil types, and cable length is determined from the GIS data. Cable resistance is determined from conductor area and cable type; for this demonstration, cables with area measured in mm² are assumed to be XLPE cables (and those measured in in² are assumed to be PILC. Current predictions are then adjusted by the factor R/L to give values for I^2R/L ; this factor is proportional to temperature, and is used as one of the inputs to the temperature model. An example of cables with this information merged in is shown in Table 9.

cable_fid	conductor_area_mm2	q95_c_transform_mu	q95_c_transform_sigma	R_per_L	area_unit	elevation	latitude	RUC	soil_thermal_resistivity	I2R_q95_mu	I2R_q95_sigma
10681993_0	95	242.328	49.733	0.000402	mm2	176.447	53.554	Urban	0.956	0.097	0.020
5808114_0	185	121.721	30.136	0.000208	mm2	183.171	53.555	Urban	0.936	0.025	0.006
5808114_1	185	188.210	40.123	0.000208	mm2	183.171	53.555	Urban	0.936	0.039	0.008
6394610_0	95	25.176	15.113	0.000402	mm2	187.375	53.555	Urban	0.936	0.010	0.006
7024950	95	367.243	47.285	0.000402	mm2	175.886	53.554	Urban	0.956	0.147	0.019
7074320_0_1	95	225.222	34.384	0.000402	mm2	175.847	53.554	Urban	0.956	0.090	0.014
7074320_1	95	141.688	30.230	0.000402	mm2	175.847	53.554	Urban	0.956	0.057	0.012
7168396_0	35	531.907	67.700	0.001083	mm2	176.276	53.554	Urban	0.956	0.576	0.073

Table 9: Table of spatial properties and predicted current values. q95_c_transform_mu is the predicted mean parameter of q95 current - the output of the fitted current model based on Equation 29, and I2R_q95_mu is this value adjusted by resistance per unit length (R_per_L).

This table has a prediction for every cable that is connected to a joint that we want to make predictions for (the current model is per-cable). Now we iterate over each joint, and select the cables that are connected to it. Of those, the cable with the largest I2R_q95_mu is selected to match to the joint — the largest adjusted current value is expected to correlate to the peak temperature in the joint. An example of cables matched for one joint is shown in Table 10, with the selected cable highlighted. The proportion of these cables that are XLPE (assumed to have conductor area units of mm²) is also recorded as this is required as a feature in the damage model later.

cable_fid	conductor_area_mm2	q95_c_transform_mu	q95_c_transform_sigma	R_per_L	area_unit	elevation	latitude	RUC	soil_thermal_resistivity	I2R_q95_mu	I2R_q95_sigma
10681993_0	95	242.328	49.733	0.000402	mm2	176.447	53.554	Urban	0.956	0.097	0.020
5808114_0	185	121.721	30.136	0.000208	mm2	183.171	53.555	Urban	0.936	0.025	0.006
5808114_1	185	188.210	40.123	0.000208	mm2	183.171	53.555	Urban	0.936	0.039	0.008
6394610_0	95	25.176	15.113	0.000402	mm2	187.375	53.555	Urban	0.936	0.010	0.006
7024950	95	367.243	47.285	0.000402	mm2	175.886	53.554	Urban	0.956	0.147	0.019
7074320_0_1	95	225.222	34.384	0.000402	mm2	175.847	53.554	Urban	0.956	0.090	0.014
7074320_1	95	141.688	30.230	0.000402	mm2	175.847	53.554	Urban	0.956	0.057	0.012
7168396_0	35	531.907	67.700	0.001083	mm2	176.276	53.554	Urban	0.956	0.576	0.073

Table 10: Table of spatial properties and predicted current values. q95_c_transform_mu is the predicted mean parameter of q95 current, and I2R_q95_mu is this value adjusted by resistance per unit length (R_per_L).

The temperature model also uses the location-based inputs of elevation, latitude, soil thermal resistivity and a rural/urban indicator as well as the adjusted current values (e.g. I2R_q95_mu and I2R_q95_sigma are inputs to the temperature q95 model). Predictions for mean and standard deviation of temperature q95, q98 and q90 range are made. The distribution of these predictions of mean value across all joints on the network is shown in Figure 78.

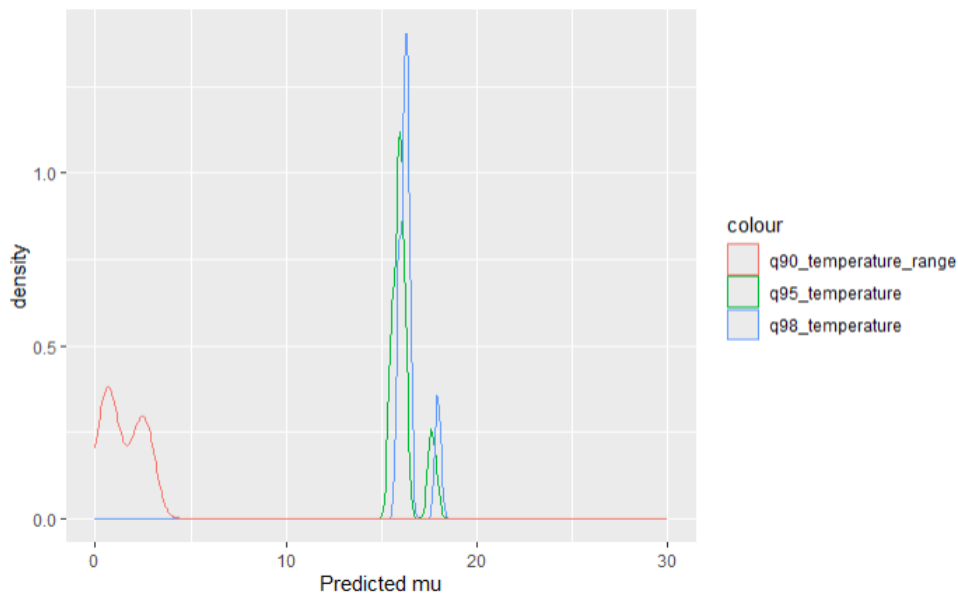


Figure 78: Distribution of predicted mean parameters for the three temperature metrics, across all joints on the network. The predicted mean parameters are based on Equation 31.

Finally, the damage model also uses the proportion of cables around the joint that are XLPE, and the predicted temperature mean and standard deviation parameters as inputs. Predictions of damage are made for each joint, and plotted on the network area map in Figure 79. While one joint appears to have slightly higher predicted damage than others, all predicted damage values are very low.

It is worth emphasising that these damage values are purely based on the outputs of the physical models with no calibration or verification against observed faults and failures. Therefore, the absolute values should not be taken as accurate, but the relative size of predicted damage at different locations may indicate differences in failure probabilities.

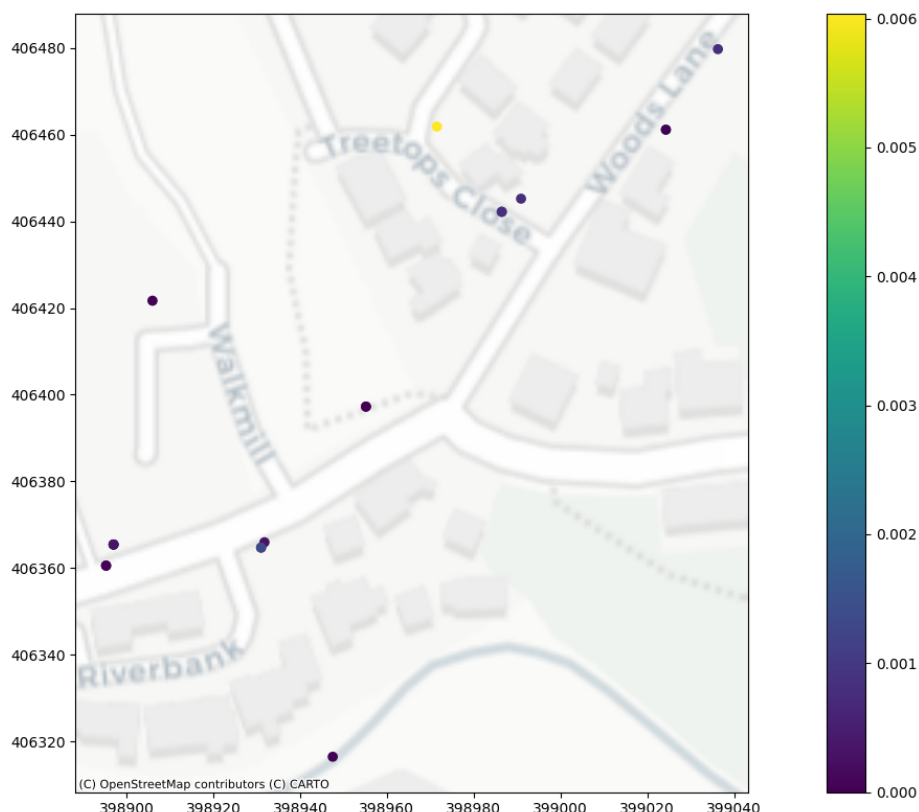


Figure 79: Map of joints with their predicted median damage values. Predictions are output of model fitted to Equation 37.

7.2 Whole-region demonstration

A demonstration of predictions for a set of new, unseen cables is now outlined and the collective properties shown. A set of cables on feeders is created from ENW's dataset of customers per service point, merged with the LV way they are connected to. To recreate examples of cables at various points on the feeder, a random subset of service points on a given LV way was selected and the number of customers and their annual demands summed to get the customer parameters for a theoretical cable - this was repeated to get an overall set of 9000 cable and joint pairs. Demands were converted to a total MWh for the year. EV and heat pump penetrations per local authority area for 2030 were taken from ENW's DFES workbook [53]. Each created cable was assigned to a local authority area randomly and the corresponding EV and heat pump penetrations merged into the cables dataset. For the purposes of this exercise, no networks were assumed to be complex.

This gave a set of cables with the distributions of customers downstream and per-customer annual energy seen in Figures 80 and 81 respectively.

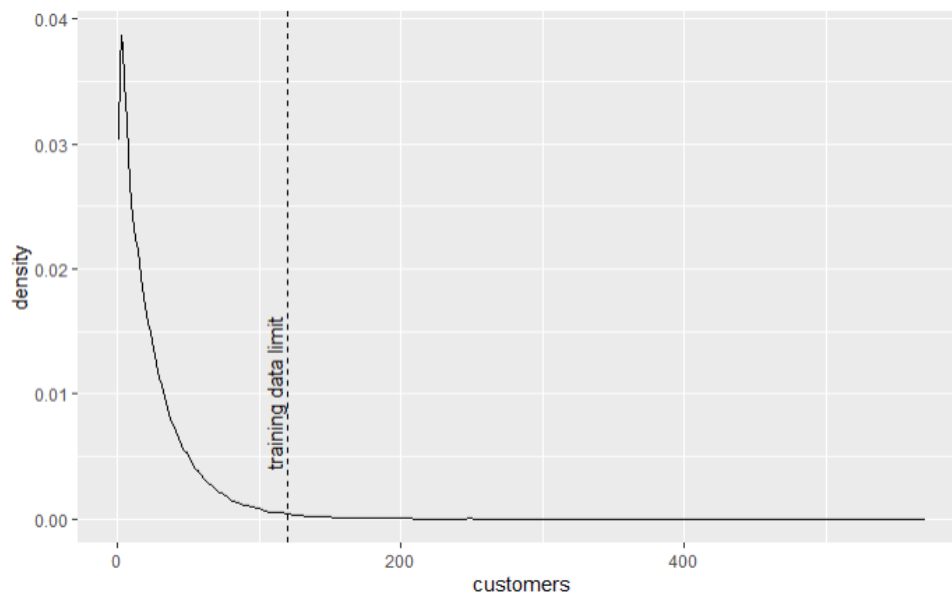


Figure 80: Distribution of number of customers downstream, across all artificially created cables.

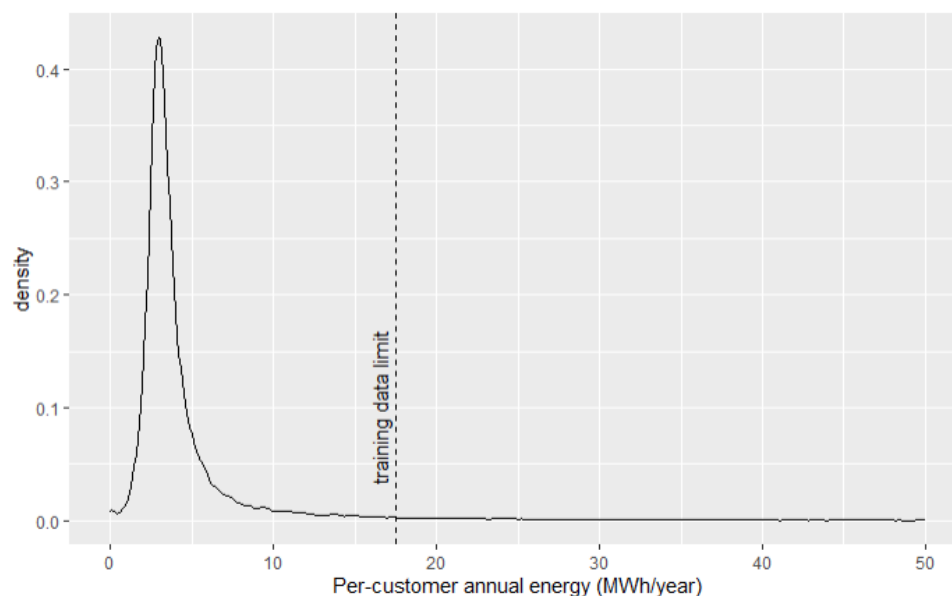


Figure 81: Distribution of per-customer annual demand, across all artificially created cables.

Fitted models for current were read in, and the new data supplied to make predictions. For current, these predictions consist of a central estimate (μ) and an estimate of uncertainty (σ) which is the central 30% probability interval as described in Section 6.1. Predictions are made for q95 and q98 of current squared, and q90 of current range from the separate fitted models for each of these target variables. The distribution of predicted central estimate and uncertainties are shown in Figures 82 and 83; these values are on the scale of the *square* of current. The distribution for the central estimate of q98 has a longer tail, as expected since $q98 > q95$.

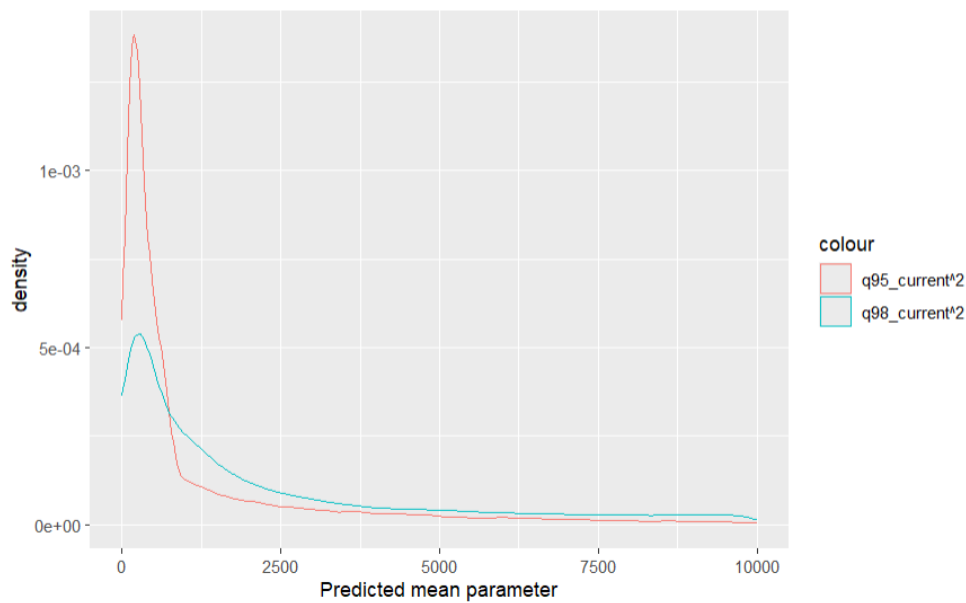


Figure 82: Distribution of predicted central estimates for q95 and q98 I^2 .

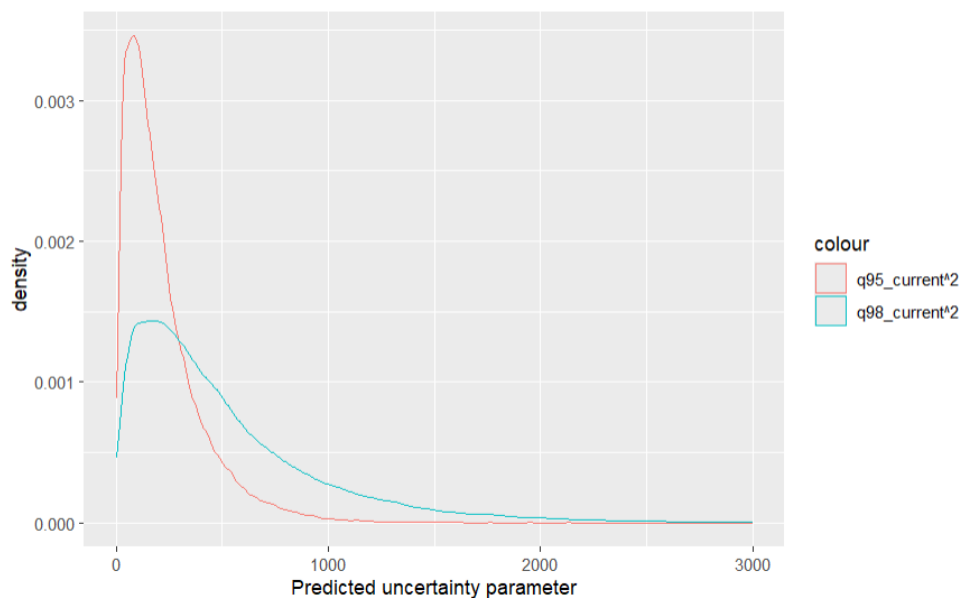


Figure 83: Distribution of predicted uncertainty estimates for q95 and q98 I^2 .

As expected, there is a positive relationship between the number of customers downstream and the predicted quantile of q95 I^2 as shown in Figure 84. Additionally, instances with higher per-customer annual demand (lighter blue points) show higher values of predicted q95 I^2 .

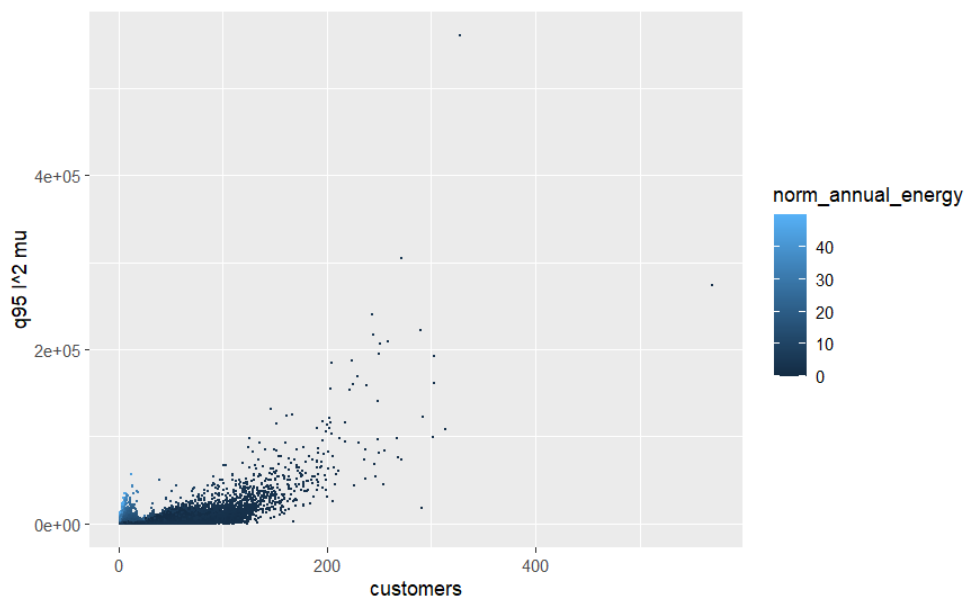


Figure 84: Number of downstream customers vs predicted central estimate of q95 I^2 .

For this demonstration on artificial cables and joints, 1:1 mapping between cables and joints is used and all cable predictions are passed to the temperature models. Temperature models are parametric, and produce predictions of the mean and standard deviation of a normal distribution. The distribution of predicted mean for all quantile targets is shown in Figure 85. While the estimated mean for daily temperature range is smaller than those for q95 and q98 of half-hourly temperature, the sigma values (an estimate of the uncertainty on the predictions) are of a similar range — below 2°C — for all three target variables (Figure 86). The distribution of predicted means for q95 and q98 temperature are both bimodal, suggesting two ‘regimes’; this effect is not seen in the distribution of current parameters so it could have been introduced by some of the additional features in the temperature model such as cable resistances or locational features like elevation or rural/urban effects. This bimodality is seen in the predictions of both the mean and uncertainty parameters.

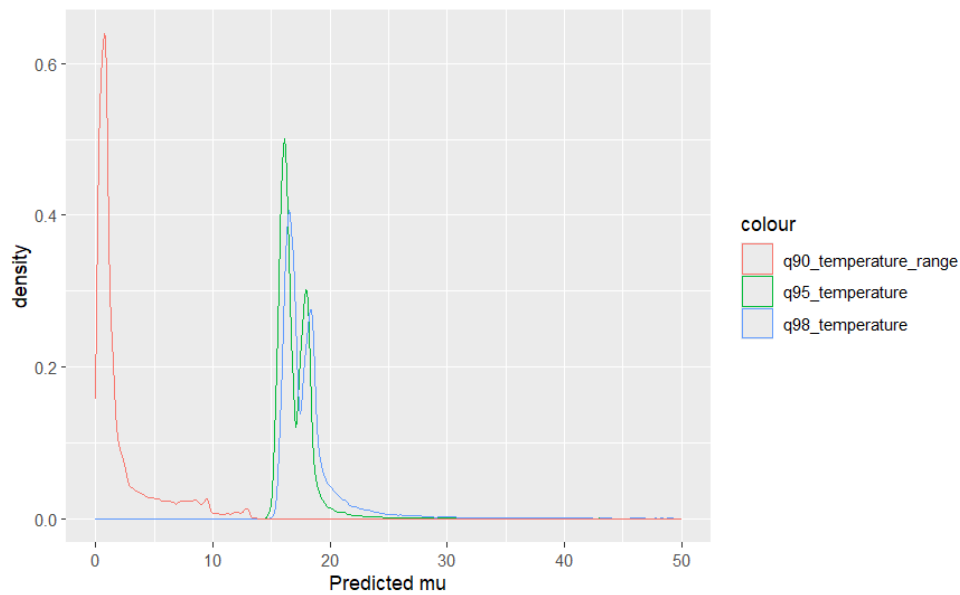


Figure 85: Distribution of predicted mean for temperature metrics.

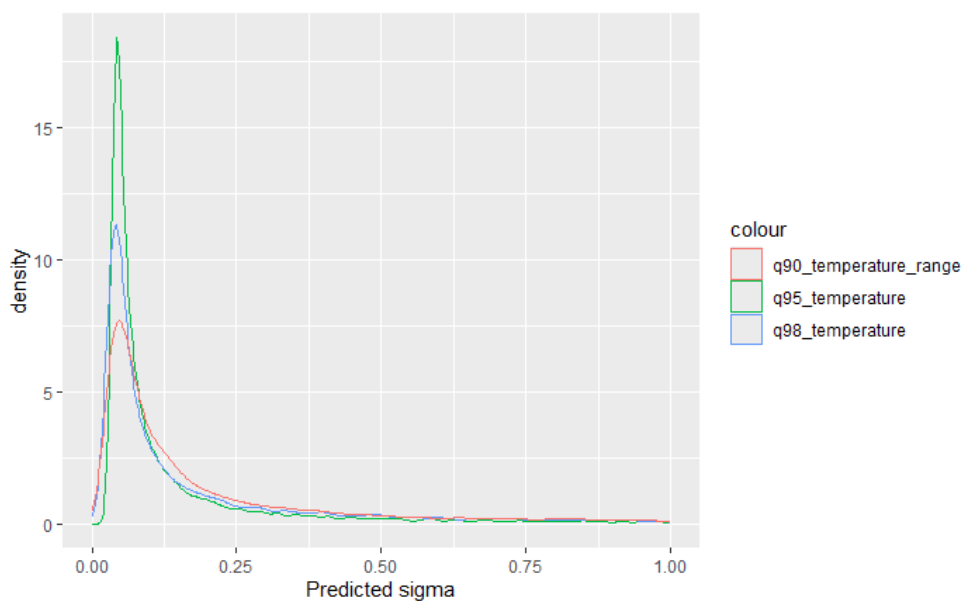


Figure 86: Distribution of predicted standard deviation for temperature metrics.

Figure 87 shows a well defined relationship between the adjusted current feature which is an input to the temperature models, and the resulting prediction of mean q95 temperature. This suggests the effect of other variables on the temperature predictions is quite small. This reinforces the hypothesis that the bimodality introduced stems from the R/L scaling factor rather than other input features.

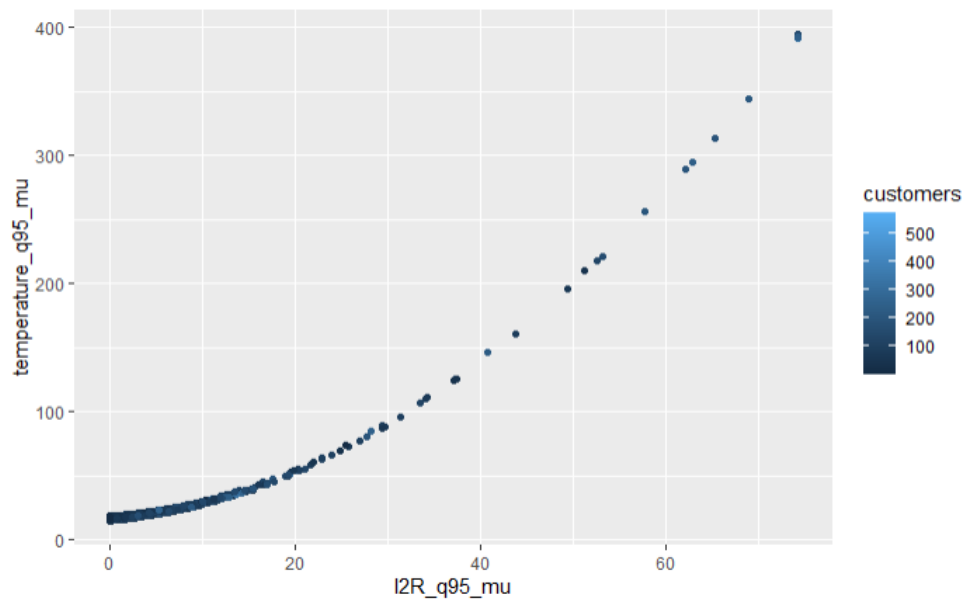


Figure 87: Relationship between input scaled current feature and predicted q95 temperature mean.

Finally, the predicted mean and standard deviation parameters of temperature are used as inputs into the damage model, to produce the final predictions for each joint. The damage model returns a set of quantiles, giving a probabilistic prediction for every joint; the 10th, 50th and 90th percentiles are plotted in Figure 88. Quantile values have been post-processed by sorting and clipping to ensure the cumulative values are correct (e.g. the predicted q60 value is always > the q55 value), and that all values are between 0 and 1.

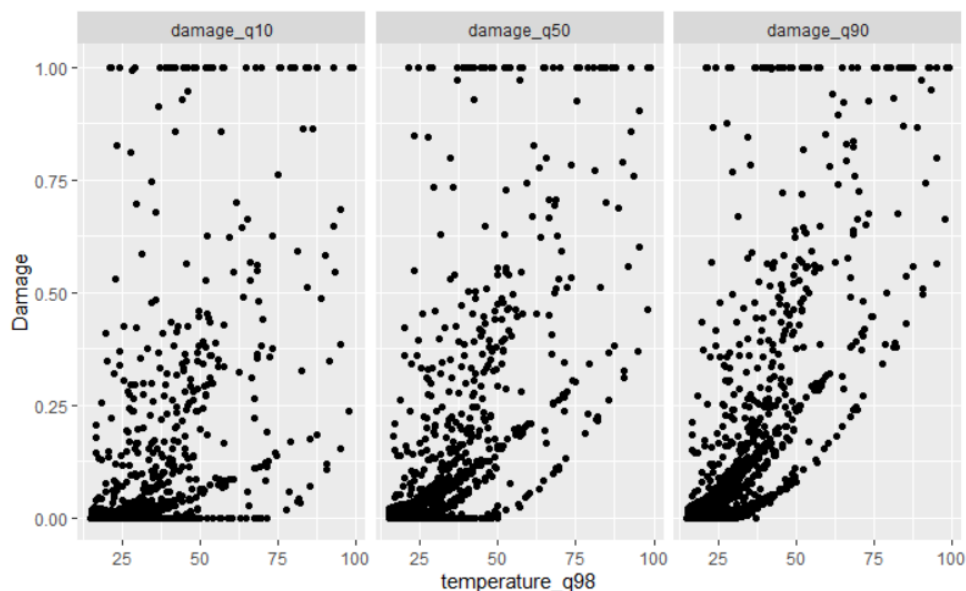


Figure 88: Relationship between input feature q98 temperature mean, and selected quantiles of the predicted damage.

These two worked examples demonstrate how the fitted statistical models enable predictions of damage for a single network and a large set of networks. The steps outlined in this section included careful use of the statistical models with consistent use of specialised packages and custom implementations. The discussion on future implementation of the statistical models is presented in Section 9.3, clarifying the risks and benefits for ENW in using the fitted models in isolation compared to implementing the physical model expressions and statistical modelling process from base data.

8 Sensitivity analyses

Assumptions in the modelling are recorded in Section 4.2. Of these, two main assumptions were identified for further analysis and investigation:

- Analysis of the potential effect of phase imbalance on network heating.
- Analysis of the impact of assuming 100% naive EV charging, compared to the likely reality of a mixture of smart and naive charging. EV demand time series of both naive and smart charging were created and full physical simulations run for both scenarios to explore the two extreme limits of all naive and all smart charging.

Both of these analyses are detailed in this chapter.

8.1 Phase imbalance sensitivity

Information about the phase connectivity of customers is sparse, with only a small proportion of customers where the phase is known. In addition, moving to models where phases are explicitly modelled is a significant increase in complexity in itself with three times the number of time series needed for each component, and an unbalanced load flow which significantly increases complexity. The complexity of the physical models for temperature and damage would also increase if they were to consider different loading and heating on different phases. Therefore, it is important to understand the effect simplifying the models to only consider balanced phases would have on the resulting modelled temperatures and damages. This sensitivity analysis aimed to establish the likely magnitude of phase imbalance seen, and the potential resulting current and temperature imbalances this could result in.

To do this, a set of 1700 cables with randomly assigned customer, EV and heat pump numbers were created. Each customer, was assigned to a phase at random, so that the total number of customers, EVs and heat pumps is on average equal across all phases, but there is variation in the exact assignments. Demands are sampled by picking time series 10 times for each cable. For every simulation and every cable, the phase imbalance ratio was calculated. This is defined as:

$$demand_ratio = \frac{\max(d_A, d_B, d_C)}{\text{mean}(d_A, d_B, d_C)} \quad (38)$$

This metric is equal to 1 when all phases are perfectly balanced, and 3 when they are completely imbalanced (i.e. all demand is on one phase). While the imbalance in demand is useful to know, cable heating is roughly proportional to the *square* of demand since temperature is proportional to I^2 and demand is proportional I . A metric for heating ratio is therefore:

$$heating_ratio = \frac{\max(d_A^2, d_B^2, d_C^2)}{\text{mean}(d_A^2, d_B^2, d_C^2)} \quad (39)$$

An example time series of the two imbalance metrics is shown in Figure 89. This shows the heating imbalance is consistently higher than the demand imbalance. The overall distribution of demand and phase imbalance across all cables and all time points is shown in Figure 90. This shows that a demand phase imbalance ratio of 2 is exceeded only 4.5% of the time, but that a heating phase imbalance ratio of 2 is exceeded 18% of the time - so even cases

that are not considered unbalanced in terms of demand could still produce quite a large imbalance in heating between the phases.

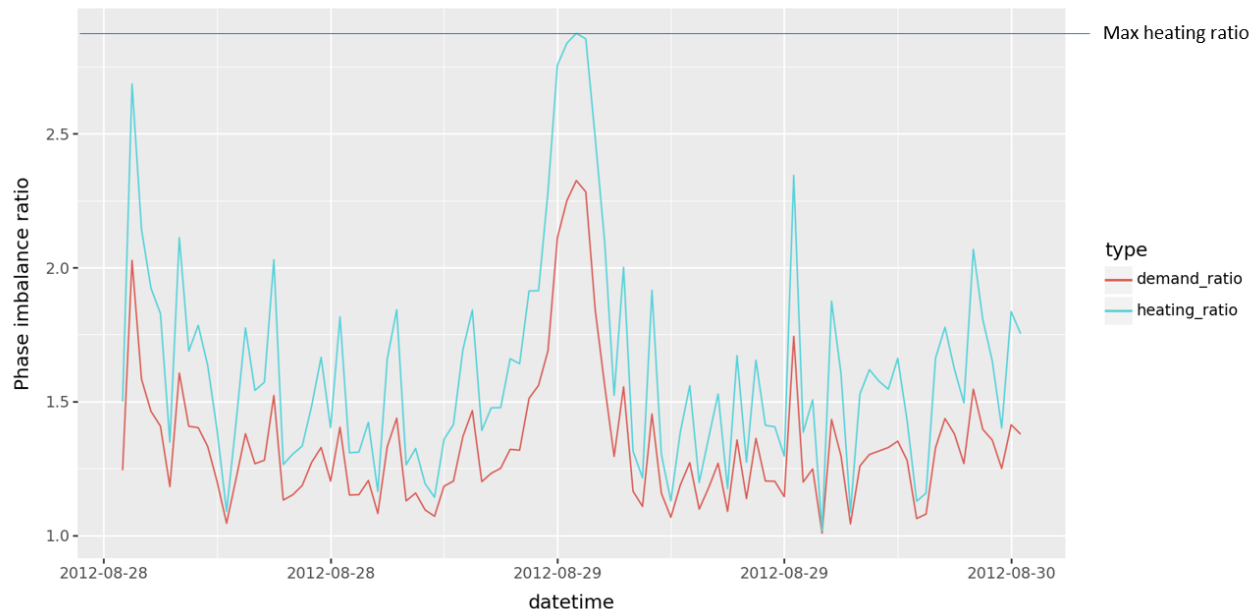


Figure 89: Example time series of demand and heating imbalance metrics.

Figure 91 shows the relationship between the number of customers on a cable and metrics of the demand phase ratio. The points in the upper left hand corner represent cables with only 1 customer — in these cases, by definition the phase ratio will have the maximum value of 3. In general, the demand phase imbalance decreases with increasing number of customers. This makes intuitive sense, since two effects serve to even out the demands across phases when the number of customers increases: firstly, shifting a single customer from one phase to another represents moving a much smaller proportion of overall demand when the number of customers is high, so having a discrepancy of say two customers between phases makes a much larger difference when the total number of customers is small. In a similar way, if one customer has unusually large demands (for example have both an EV and heat pump while other customers have neither), this will again represent a much smaller proportional increase on one phase when they are part of a large group of customers. The same relationships are seen for heating imbalance in 92, but with consistently higher phase ratios across all customer numbers, but with an even larger increase for lower customer numbers.

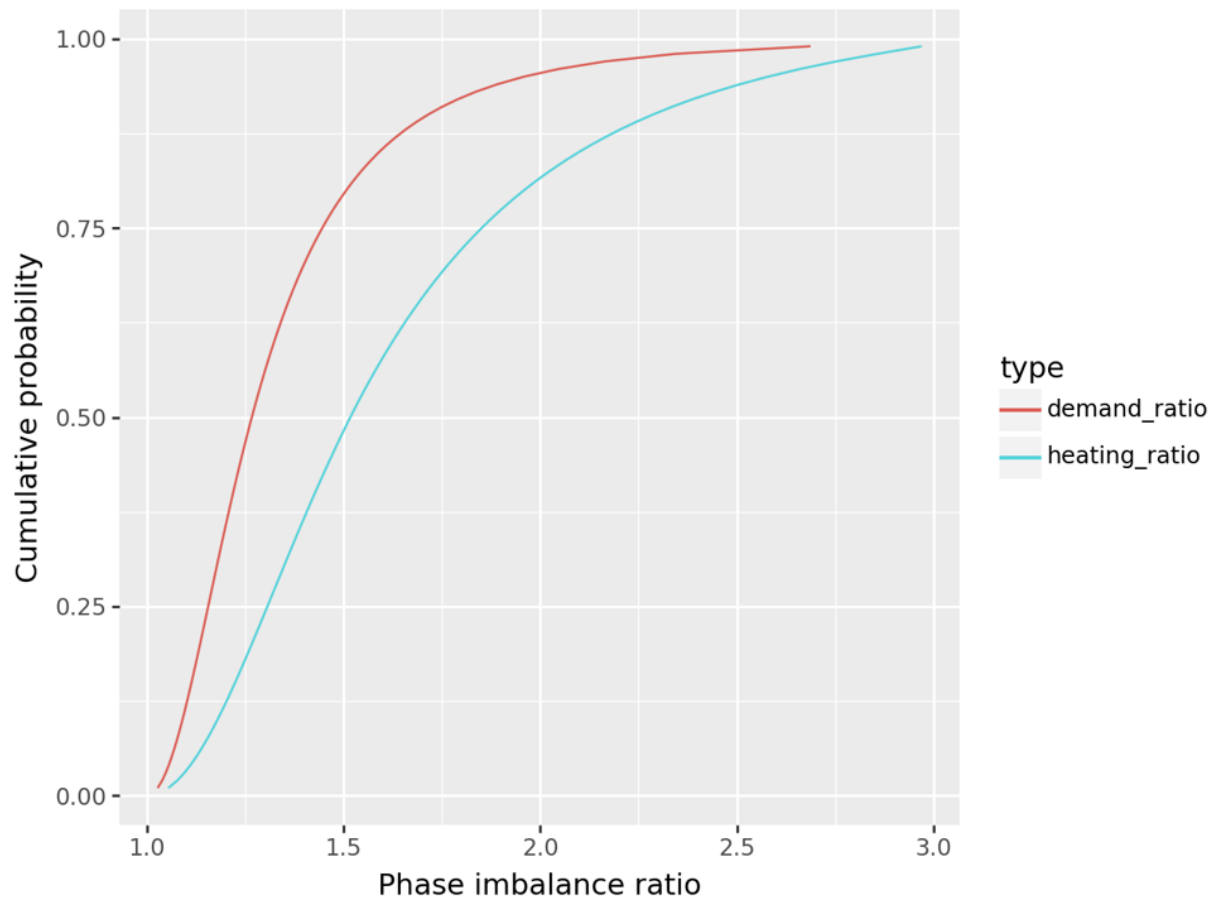


Figure 90: Cumulative distributions of demand and heating phase imbalance metrics, across all cables and all time points.

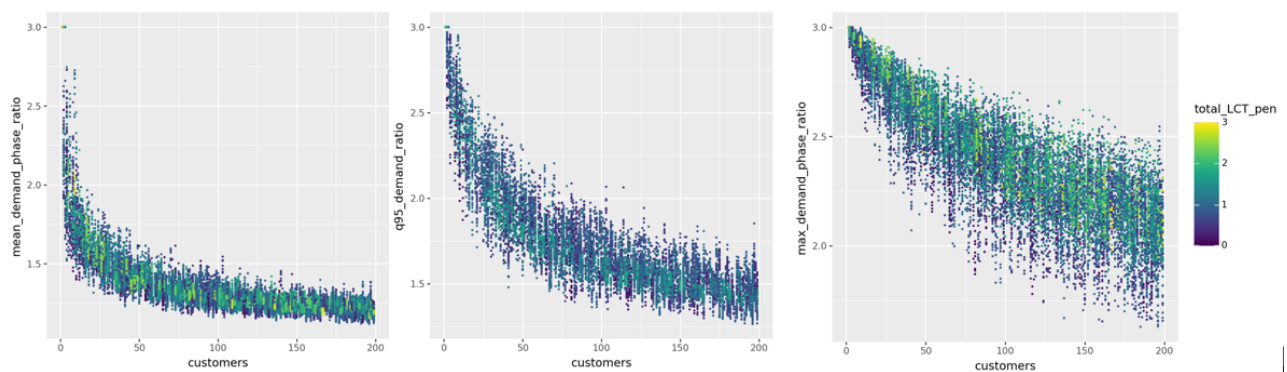


Figure 91: Relationship between the number of customers on a cable and metrics of the demand phase ratio - the mean across all time points (left), q95 (centre) and maximum (right).

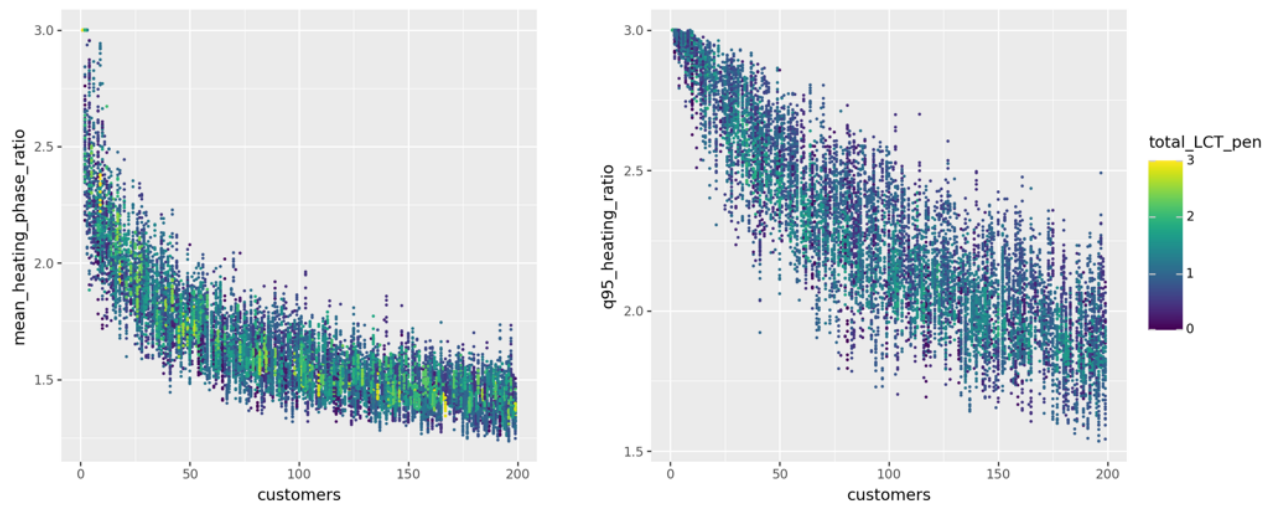


Figure 92: Relationship between the number of customers on a cable and metrics of the heating phase ratio - the mean across all time points (left), q95 (right).

There is no clear relationship to the number of installed LCTs discernable from the colouring on Figures 91 and 92; however, there is some relationship between the imbalance of LCTs across phases and the overall demand and heating phase imbalances as seen in Figure 93. In particular, this plot shows an increase in the lower limit for mean phase ratio when the penetration of LCTs is large.

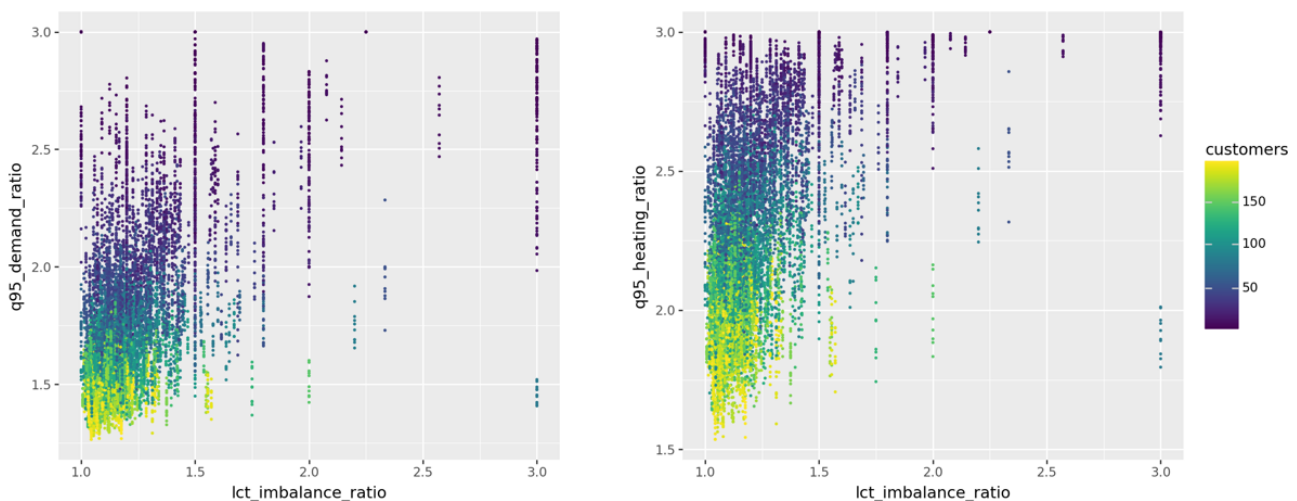


Figure 93: Relationship between the imbalance of LCT numbers between phases, and the mean demand (left) and heating (right) phase ratios.

The overall strength of correlation between the various cable features and the resulting phase imbalance is shown in Table 11.

Overall, this sensitivity analysis has shown that, in general, quite large demand and heating imbalances can result from randomly (and therefore generally fairly equal) assignments of customers to phases. Heating phase imbalance — crucial for temperatures and therefore degradation in the cables and joints — is consistently greater than demand

Variable	R^2 against q95 demand phase imbalance	R^2 against q95 heating phase imbalance
# customers	0.67	0.76
# EVs	0.43	0.57
# heat pumps	0.18	0.17
LCT phase imbalance	0.29	0.21

Table 11: Strength of relationship between cable features and phase imbalances.

phase imbalance. This is not something that has received significant prior research focus but should be considered in future work.

Phase imbalance is more strongly related to the number of customers than the LCT penetration, although LCT imbalance across the phases is more strongly related to demand and heating phase imbalances than the total number of LCTs present. Lower customer numbers are more likely to result in high phase imbalances, due to the effect of discrete small numbers meaning a discrepancy of a single customer between phases has a much larger effect on the phase imbalance than for larger groups of customers.

This work has shown that phase imbalance is a significant effect and high heating imbalances between phases are likely in practice. Therefore, it would be important to consider phase imbalance in the physical modelling. However, the practical complexity of this was too great to feasibly be implemented within this project, as it would not only significantly increase the complexity and number of modelled time series required, but also require an extra 'dimension' of sampling, increasing the number of samples needed for good coverage by orders of magnitude. There was no simplified representation of possible phase imbalance or rules of thumb that could have given an initial rough quantification of the possible effect of phase imbalance on temperature and damage, and so without more complete data on actual phase allocations, it was decided to exclude any considerations of phase imbalance from the physical modelling with the acknowledgement that this is an important factor that would likely increase damage beyond that seen for the modelled balanced scenarios.

8.2 EV charging behaviour sensitivity

The effect of all EVs using naive charging vs all EVs using smart charging was investigated. The same set of physical simulations for the same networks were run with both the naive and smart charging EV time series as described in Section 4.1.7.2, and here the results of those simulations are compared. Figure 28 shows that while smart charging shifts EV demand to the low-demand overnight period, the peak demand from a group of EVs is much higher since they all start charging basically simultaneously at the start of the off-peak period. A different charging regime that spreads the total demand over the entire off-peak period may help with this but may be less likely to be widely adopted than the simple dual tariff modelled here.

The set of cables with both smart and naive charging simulations is determined, and the datasets joined for comparison. Given the two datasets have the same simulations of the same cables, with the same number of customers, EVs and heat pumps it is not surprising that Figure 94 shows the distribution of total annual energy is almost identical:

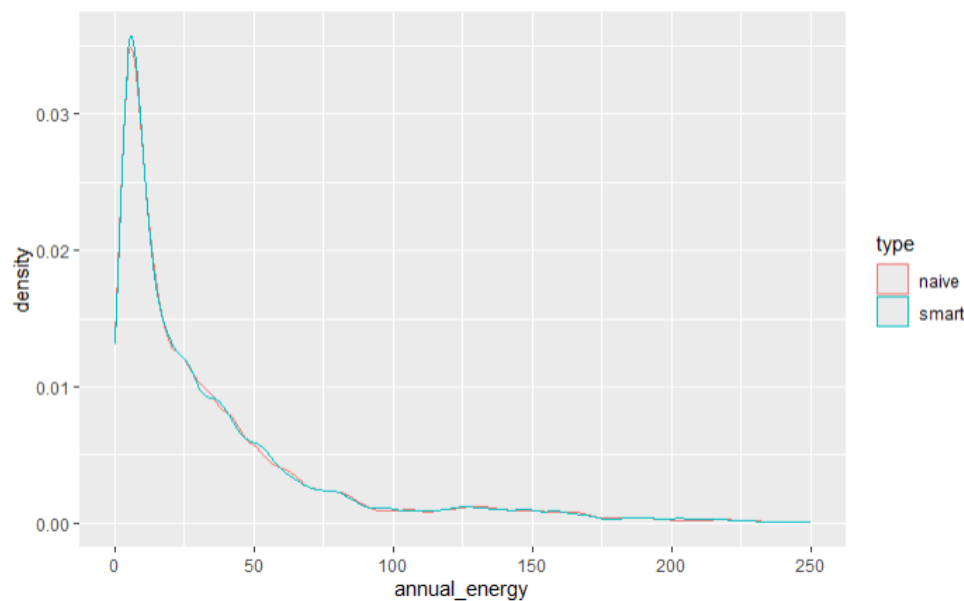


Figure 94: Distribution of total annual energy for sets of cables, with both naive and smart charged EVs.

The distribution of high quantiles of current may well differ between the two datasets, as the timing of EV demands is different. As Figure 95 shows, there are differences between naive and smart charged EVs. In particular, there is a greater proportion of cases with either the lowest or highest q95 currents in the smart-charged EV scenario. The increase in low values of q95 current for smart charged EVs may occur when shifting the EV to overnight tends to create two smaller peaks (in the evening with the remaining demand, and overnight with EV demand). By contrast, if there are a particularly large number of EVs downstream and they all start charging simultaneously overnight, this new overnight peak may be greater than the peak in the naive charging series where the exact timing of EV plug-ins is more dispersed, therefore slightly smoothing and flattening their effect on peak demand.

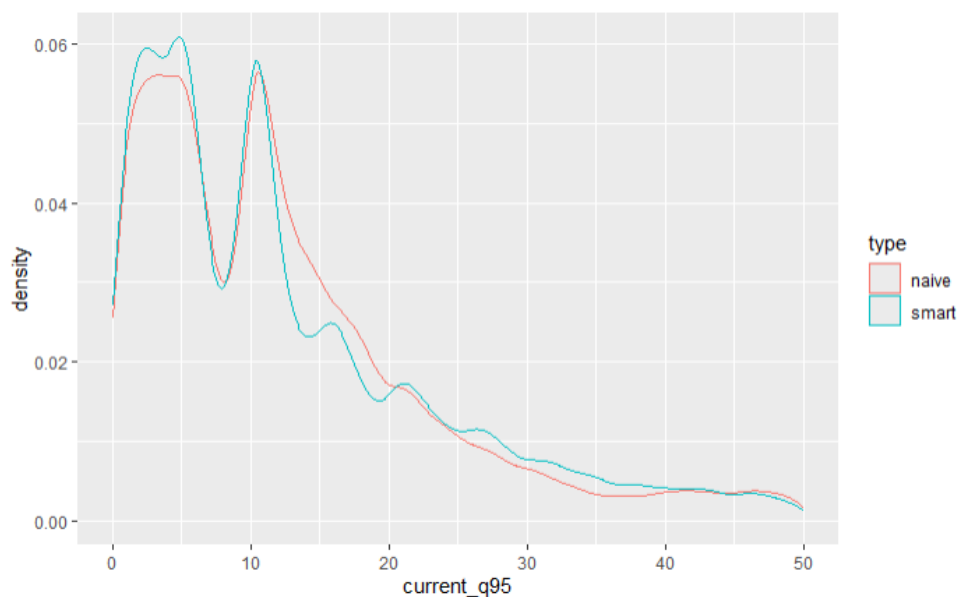


Figure 95: Distribution of q95 current for sets of cables, with both naive and smart charged EVs.

This hypothesis is backed up by the relative changes in distribution when the dataset is split by EV penetration (Figure 96): here the distribution of q95 current is almost identical when the EV penetration is below 0.5, but shifts to higher currents for smart charging EVs at higher EV penetrations. The same phenomenon is displayed in the distribution of q90 of daily current range in Figure 97, showing that the daily variability in current also increases under the smart EV charging regime for high EV penetrations.

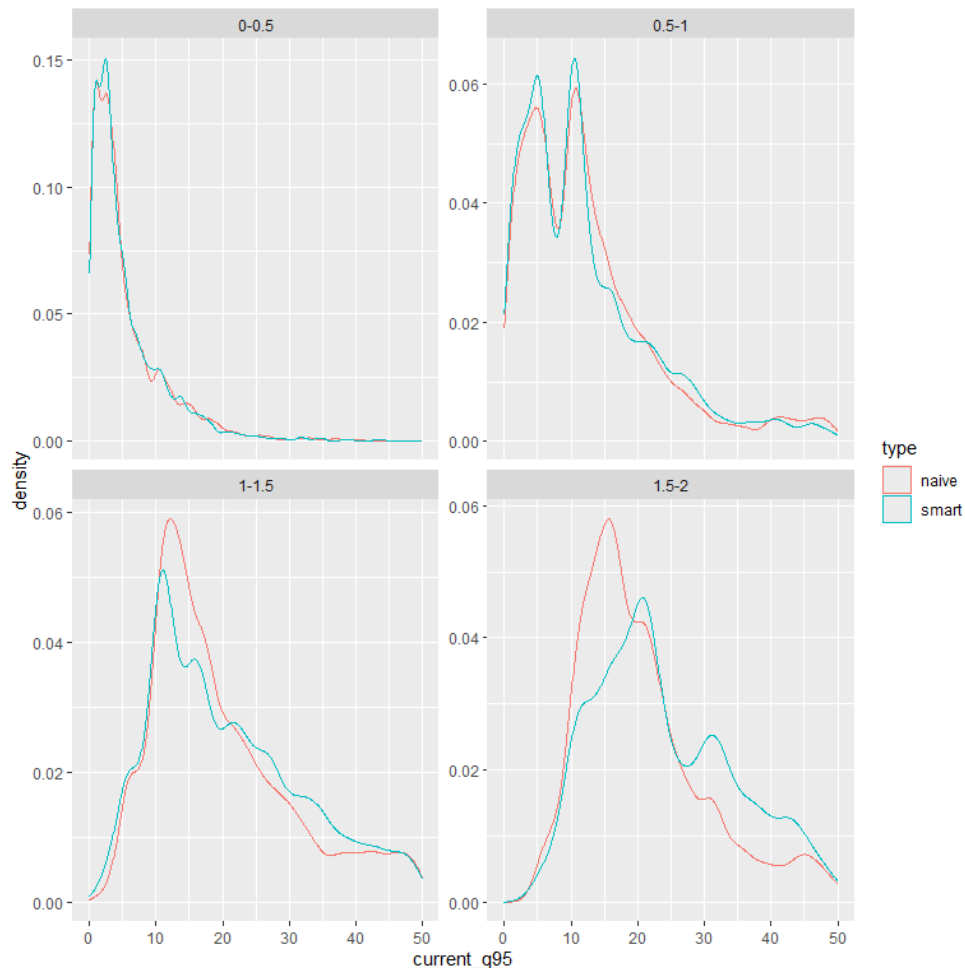


Figure 96: Distribution of q95 current for sets of cables, with both naive and smart charged EVs. Facets show different subgroups of EV penetration.

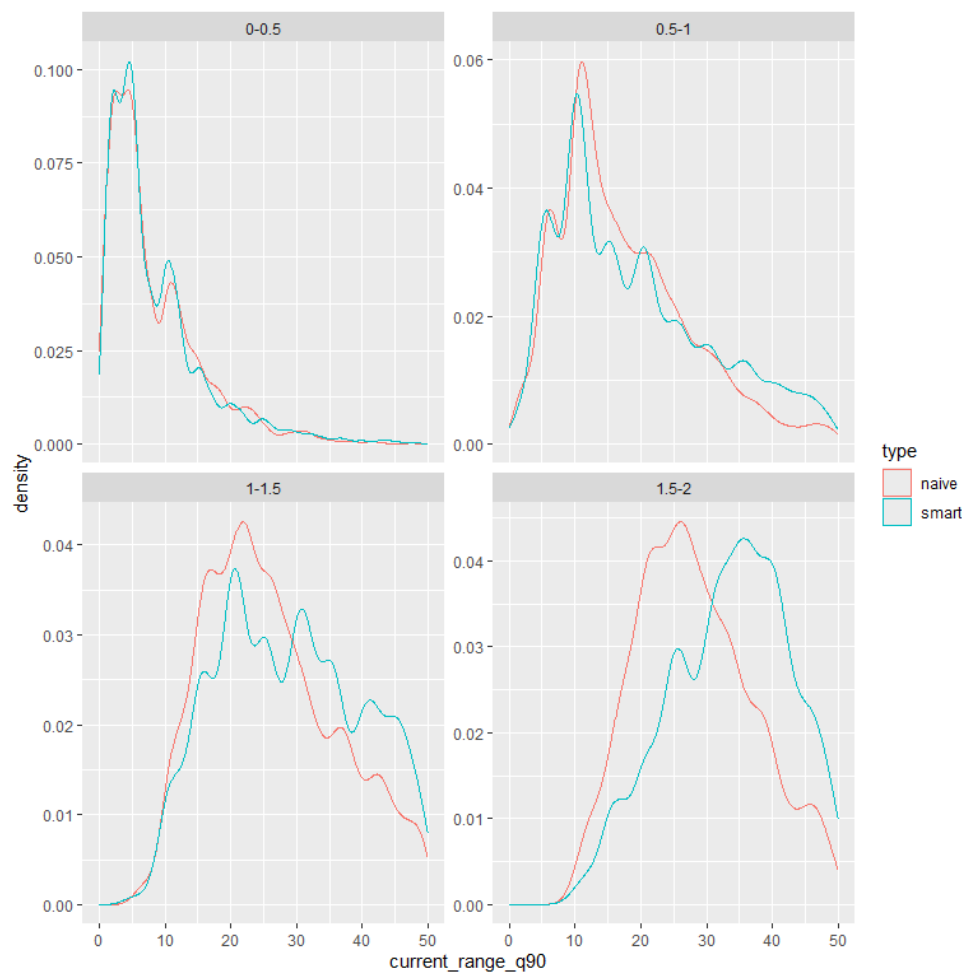


Figure 97: Distribution of q90 current range for sets of cables, with both naive and smart charged EVs. Facets show different subgroups of EV penetration.

Similar analysis was undertaken for temperature and damage metrics, although these were measured per joint rather than per cable. Joints were matched between the naive and smart charging datasets in the same way as for the set of cable results that recorded cable demand and current. Interestingly, the difference between naive and smart EV charging at high EV penetrations does not appear to affect the distribution of high temperatures seen: the subsetting distributions of both q95 and q98 temperature are very similar between naive and smart charged EVs. This is shown in Figure 98, with only a very slight increase in q98 of temperature for the highest category of EV penetration under smart charging.

The change in charging type has more of an effect on the daily range of temperature, and this change is in line with that seen in daily current range above.

Finally, the effect of different EV charging regimes on damage is analysed in Figure 100. Note the scale: the vast majority of damage values are very small, and the distribution has a long tail as seen by the box and whisker plots with outliers in Figure 101. While the main distribution of damage shifts higher for high penetrations of smart charged EVs, the number of instances of much larger values of damage also increases.

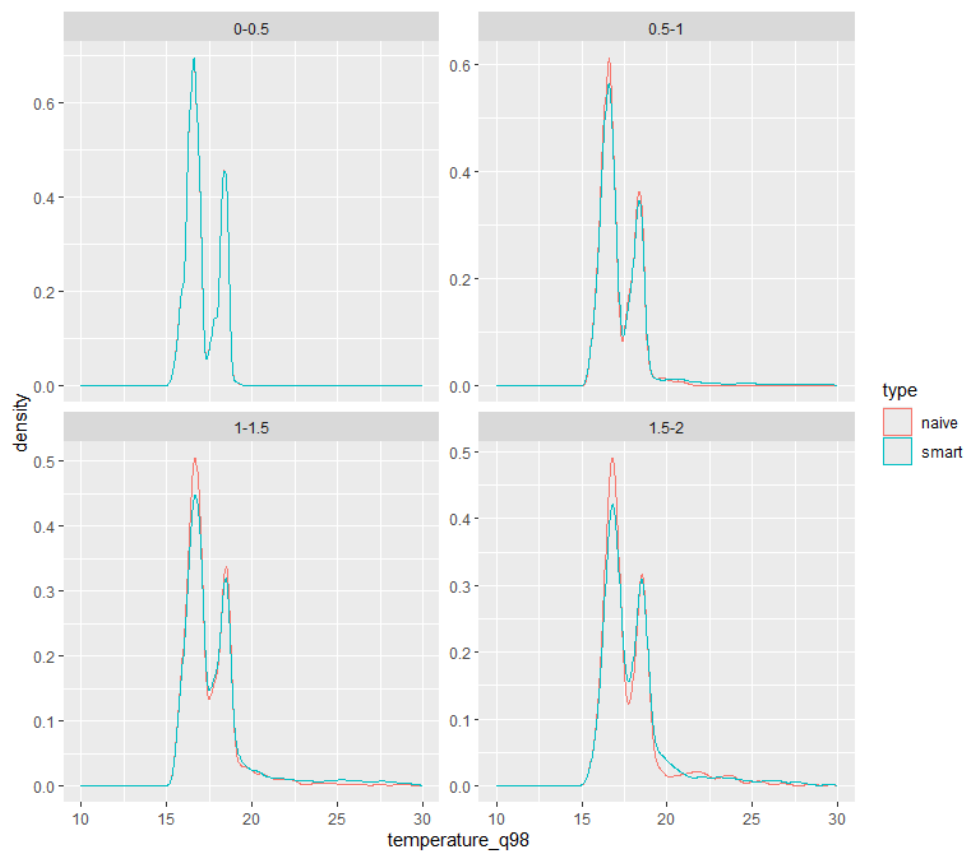


Figure 98: Distribution of q98 temperature for sets of joints, with both naive and smart charged EVs. Facets show different subgroups of EV penetration.

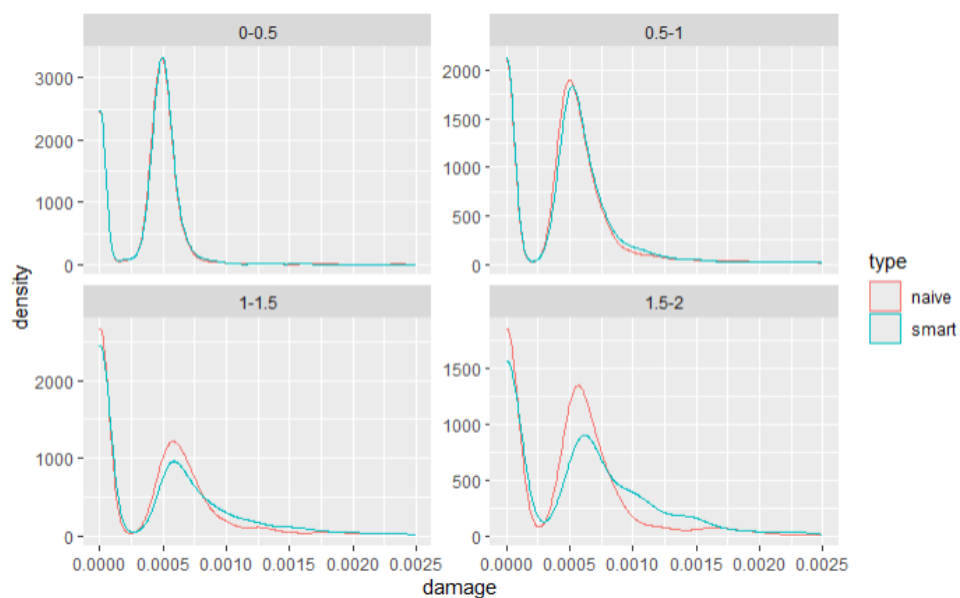


Figure 100: Distribution of damage for sets of joints, with both naive and smart charged EVs. Facets show different subgroups of EV penetration.

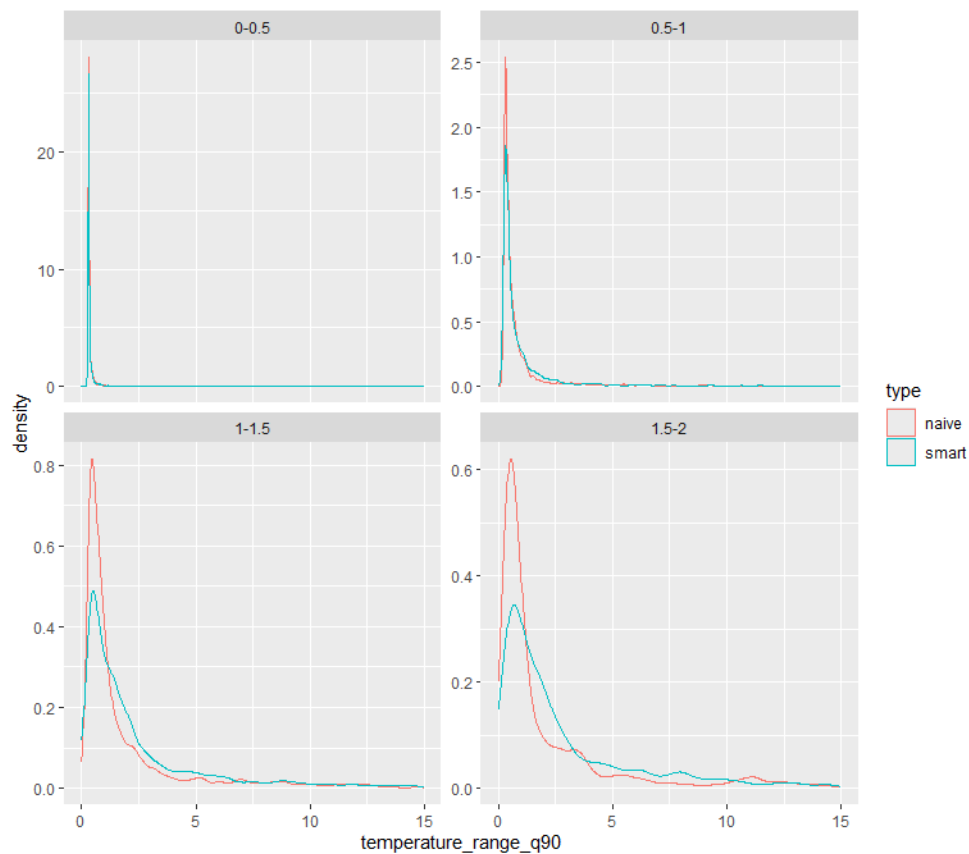


Figure 99: Distribution of q90 temperature range for sets of joints, with both naive and smart charged EVs. Facets show different subgroups of EV penetration.

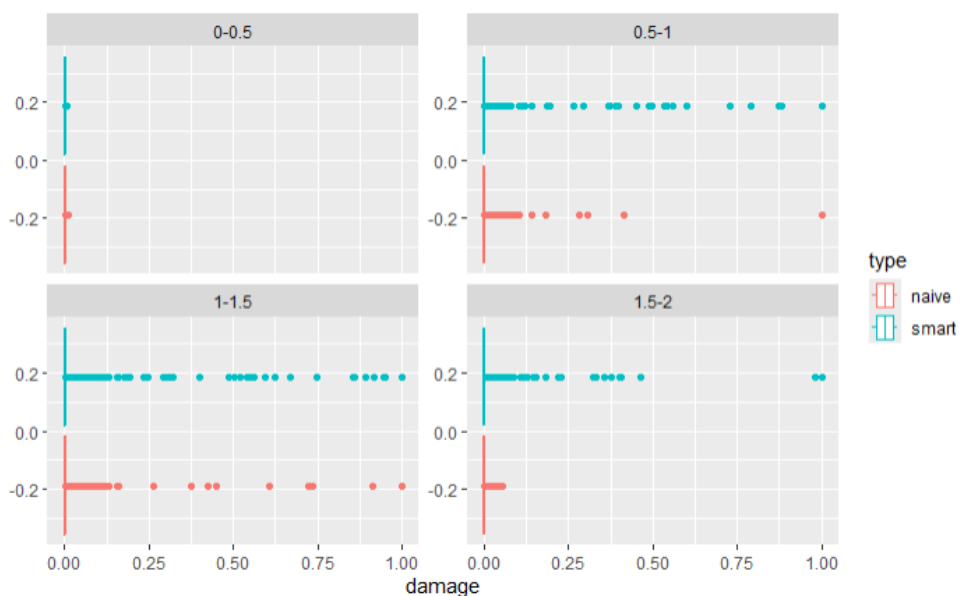


Figure 101: Box and whisker diagrams, with outliers, of damage for sets of joints, with both naive and smart charged EVs. Facets show different subgroups of EV penetration.

Smart charging might have been expected to reduce the highest demands, currents, temperature and damages seen on networks by shifting EV demand to off peak periods. However, this effect appears limited at low EV penetrations, and the effect of all EVs simultaneously charging at the start of the off-peak period negates any positive effect of this shift in time for larger EV penetrations. Smoothing the total EV demand over the entire off-peak period, and thereby reducing the maximum demand drawn at any given time for the EV, would likely mitigate this effect but would require different charging systems and knowledge of the car's starting and end state of charge in order to set the charging demand level. While this is technologically possible, there is potentially less certainty over this becoming widespread and the norm charging behaviour and so was not modelled in this project.

9 Discussion

LV Predict 2 has built on the initial methodology demonstrated as a proof-of-concept in LV Predict 1, following the same fundamental physical processes of demand-driven cable heating and degradation. Reviewing the simplifying assumptions made in LV Predict 1 and refining the method for more realistic modelling has resulted in improved understanding of the cable conditions and refinements to the modelling, which are summarised here.

9.1 Physical models

Returning to the physical modelling assumptions from LV Predict 1 with more research time and additional expertise has benefited the degradation prediction outputs. This includes returning to the underlying FEA modelling, which demonstrated a different stress profile in the joint under thermal loads, as the effect of temperature-gradient induced stresses is accounted for alongside differential thermal expansion of a cable's layers. For all three damage mechanisms that utilise the stress model, the factor converting from temperatures to stress has reduced. This is primarily driven by the removal of a postulated defect (an initial crack) in the stress models; a key site for stress concentrations, but whose position in any given joint, if any, is very uncertain. The new stress models are judged to be more robustly applicable to the LV joint population, as the stress distribution is not skewed towards this uncertain postulated defect.

One of the key findings of the LV Predict 1 physical model investigations was that time at high temperatures was the dominant driver of degradation, above thermal cycles by several orders of magnitude in a sensitivity analysis (Section 4.2.3 in LV Predict 1 [9]). This dominance may be reduced for the updated LV Predict 2 physical model, due to the consideration of transient temperatures. Figure 102 shows how the area under the steady-state and transient temperature histories (which can be thought of as an indicator of some damage mechanisms (creep and chemical)) for the example profile in Figure 3 differs. The transient model reaches a lower temperature peak more slowly, but also cools down more gradually after demand reduces back down. Note however that the number of cycles is unchanged, indicating a different effect on the cycle-based damage mechanisms (fatigue and accumulation of plastic strains).

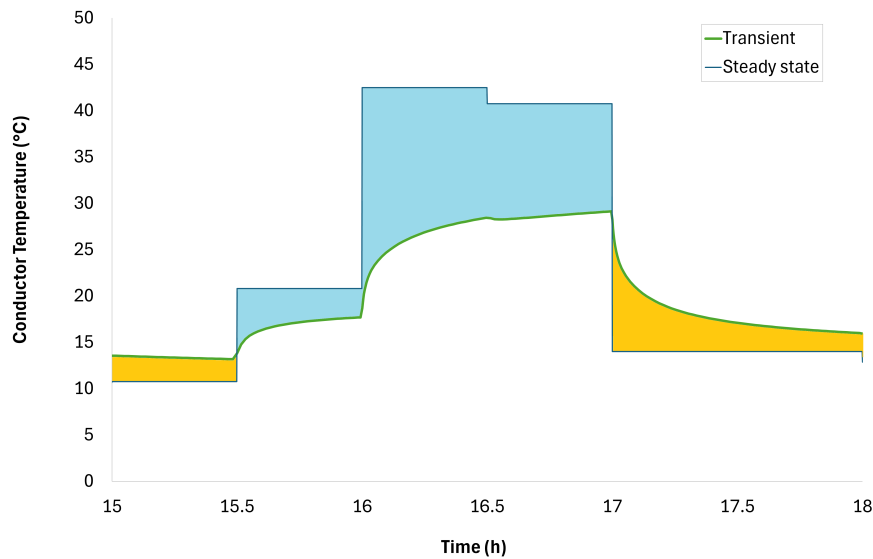


Figure 102: Extract of Figure 3 showing the area between steady-state and transient temperature predictions.

Trials on the physical models using some test properties indicated different behaviours at extremes of temperature, and this was reflected in the findings of the statistical model (discussed in 9.2). This links to the different activation energies used in several of the physical model formulae which use Arrhenius laws (creep and chemical damage). In the Arrhenius law, the increase in rate of a process is governed by an activation energy for the process via an exponential relationship (see Equation 17)). The two chemical degradation formulae and the creep formulae for the aluminium connector bolts all had significantly higher activation energies than the XLPE formula. In the temperature range of interest, this corresponds to a much steeper increase in process rate with increased temperature. This is demonstrated by differentiating the Arrhenius rate expression with respect to temperature - see Equation 40. For the relatively low temperature values experienced by the LV cables, a higher Q term in $\frac{AQ}{RT^2}$ dominates the comparison between the four Arrhenius-based mechanisms in the physical model.

$$\frac{d(\text{Rate})}{dT} = \frac{AQ}{RT^2} \exp \frac{-Q}{RT} \quad (40)$$

Figure 103 compares the Arrhenius law rate terms, normalised against their value at room temperature. The same trends are plotted using a logarithmic scale for the y-axis, to demonstrate the differences between activation energies and the magnitude of rates at higher temperatures.

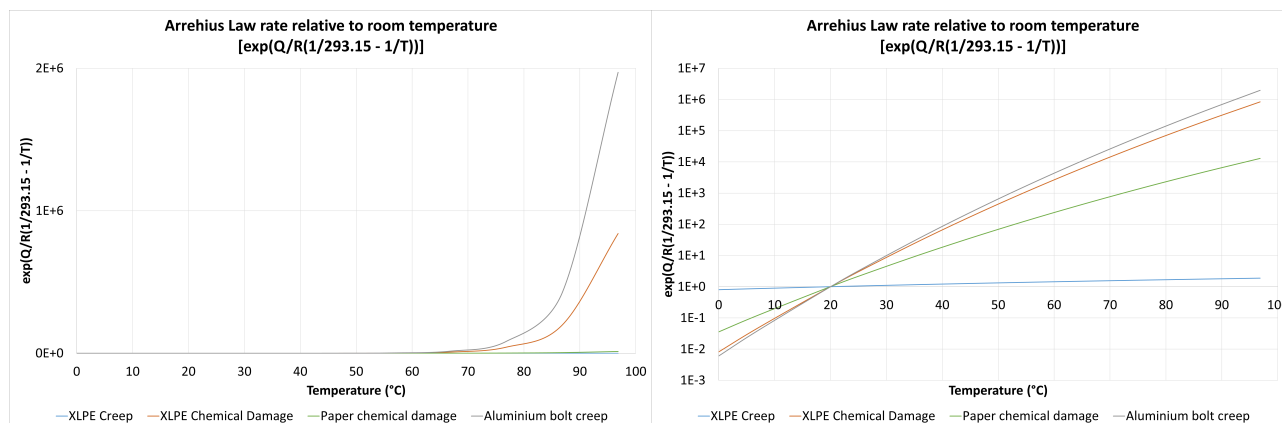


Figure 103: Comparison of process rates for the mechanisms using Arrhenius laws.

These material properties indicate that joints in the network running mostly at moderate temperatures ($\sim 30^{\circ}\text{C}$) may see very little accumulation of damage other than of creep strain in the XLPE insulation (if it is present in the joint); however even short spikes of high temperatures may incur significant damage if cables go above around 70°C . This potentially highlights the risk of an ageing network where buried cables see demands well beyond their design.

9.2 Statistical models

The findings of the statistical models were linked to the changes to the physical models, as developments in the modelling of temperature and damage are reflected in the relationships seen in the training data for the statistical models.

The addition of load flow modelling to determine cable currents from simulated demands meant LV Predict 2 introduced network topology as an additional influencing factor in the modelled currents. For the majority (96%) of networks, skilful statistical predictions of currents were still possible from only basic features with no knowledge of network topology beyond the total feeder size. However, a particular type of 'complex' network was also identified where the skill of current predictions was much more variable across the network. These networks typically had a large imbalance⁸ of customers between LV feeders, with a large number of customers on one feeder. The inclusion of input features describing network topology in a more detailed way (for example impedances between key points) may add skill to the statistical model for current for these networks; alternatively, full simulations could be carried out for this subset of networks given the relatively small size of the group.

Statistical models for temperature focussed on predicting high quantiles of temperature, and high quantiles of daily variability. The q95 temperature values seen in the training data mainly fell in the range $0-26.8^{\circ}\text{C}$, with only 1% of values of q95 temperature above this - likely a result of the effect of transient (as opposed to static) temperatures reducing the time joints experience the most extreme high temperatures. Similarly, q90 of the daily range in temperature only exceeded 27°C 1% of the time, meaning cases of joints with high sustained temperatures or large temperature cycles made up a very small proportion of the statistical model training dataset. On the one hand, this suggests high temperatures are relatively uncommon under the modelled demand and network conditions but this does reduce the statistical model's ability to accurately predict cases of high temperature when they do occur.

⁸ note this is an imbalance of customer numbers between LV ways, *not* a phase imbalance.

As for the current models, the statistical models for temperature show good skill for the majority of joints but are significantly less skilful for joints on the ‘complex’ networks - this is likely due to the carrying forward of error from the current predictions.

LV Predict 2 expanded the set of cable and joint types modelled; the specific mix of cable types connecting to the joint was found to influence how temperature affects the accumulation of damage (the different groups of coloured points in Figure 104) — in particular, joints of only PILC cables accumulate much less damage for the same experienced temperature quantiles. This is due to the types of damage modelled for each type of joint; these PILC-only joints are assumed to have bituminous filler rather than polymer filler so creep, plastic and fatigue damage are ignored meaning the only contribution to damage for these joints is chemical damage in the cable paper insulation. This effect was accounted for by inclusion of a feature specifying the proportion of XLPE cables at the joint as an input to the statistical model for damage.

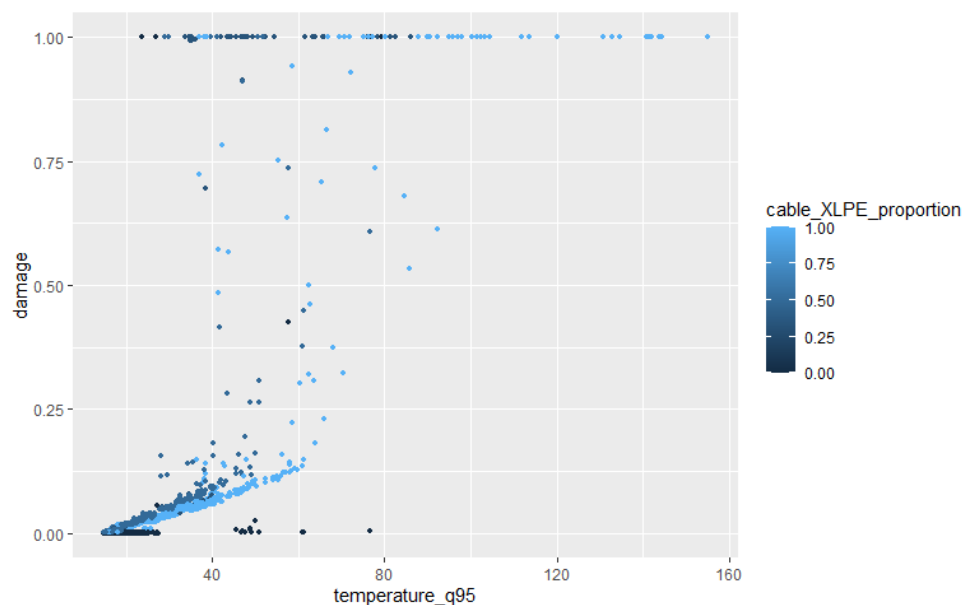


Figure 104: Relationship between simulated q95 temperature and damage.

While the relationships between *simulated* temperature quantiles and damage appear well defined, *predicted* temperature quantiles are much less skilful in predicting damage. This suggests two things: one, that despite additional complexity in the physical model for damage, there is still a strong relationship between the proposed metrics of temperature and damage — and therefore these metrics of temperature remain appropriate. However, it also suggests that while the overall errors of the temperature models are relatively low, the imperfect predictions carried forwards as input features to the damage model are nevertheless reducing the ability to distinguish cases of high and low damage with skill. The skill of the temperature models could be improved by considering more informative features of current, such as modelling more specific features of current cycling and prolonged high currents, e.g. a high quantile of the 4-hour average current.

9.3 Implementation considerations

Once the statistical models are fitted, it is theoretically possible to make predictions purely from these fitted models without the need for ENW to implement the physical model expressions and statistical modelling methodologies described in this report.

However, there are practical reasons why this is not recommended, and also benefits to implementing the methodology to fit from base data:

- ▶ Although they reduce the input data and processing and modelling capabilities required, the fitted statistical models remain relatively complex. They consist of many parameters and require familiarity with the associated packages and the origins of the mathematics to ensure consistent use:
 - The statistical models are based on GAMs, which in themselves have a more complex set of coefficients (associated with basis functions that make up splines).
 - Two-layer models (i.e. central and uncertainty models) are used. These have been created using specialised packages and customised implementations.

Given this complexity, the risk of erroneous predictions from using the fitted models as a 'black box' is considered high. For example, where link functions (a type of transformation of the target variable within the fitted model) have been used, care must be taken to ensure the predictions are returned on the scale of the original variable and not on the transformed scale.

- ▶ Relying on static fitted models does not allow for future extension, adaptation, or improvement of the modelling process. While the base training dataset has been developed to take into account future changes in demand (in particular, uptake of LCTs) this has relied on relatively old demand datasets and modelled a small subset of LV networks, as noted in Section 4.1.7.1. Likewise, refitting from base data would be required to adapt the fitted parameters for a different DNO network.

Implementing the network modelling and physical modelling of cable and joint temperature and damage would allow DNOs the flexibility to make these improvements, periodically update the model, and account for any new or unusual network topologies in future. It would also allow extraction of new features at different time resolutions, that may help improve the skill of the statistical models, in particular the temperature predictions that reflect dynamic effects.

- ▶ As discussed in Section 6.3.3, the modelling found a small subset of networks, labelled 'complex' networks, where statistical models provided significantly less skilful predictions. It is recommended that the most effective approach may be to model these types of networks with full simulations, which would require implementation of the physical modelling and simulation methodologies.

One of the key findings of this work is that there are no easily formulated, simplified representations of the complex physical processes that lead to cable degradation and damage across a network. Thus, in order for ENW to implement a robust and maintainable prediction tool for joint failures within their business, the full modelling capabilities presented in this project would be needed. As noted in earlier Chapters, assumptions and limitations remain in the modelling that would merit addressing while developing a business-as-usual prediction tool.

However, there is significant benefit from this project's learnings and qualitative outputs without a full technical implementation. The progress made in understanding the relationships between customer numbers, network topology, cable temperature, and joint damage can inform maintenance decisions and work focus. Even very simple

prioritisation metrics accounting for the impact of a fault, such as number of customers that would be affected and projected cost, could be utilised to improve decision-making with a net financial benefit.

This project's Cost-Benefit Analysis (CBA) methodology report [1] gives more detail on the benefits, both financial and non-financial, of adopting a proactive maintenance strategy that would rely on predictions of failure rates as an input. This CBA methods work has investigated the level of skill in targeting proactive maintenance that would be needed to achieve a significant level of benefit. The findings from this can inform discussions on the level of investment in in-house modelling capability that would be worthwhile to implement.

10 Conclusions and recommendations

The research sprints of the LV Predict 2 project have enabled major progress on both physical and statistical modelling compared to the previous PoC, and informed suggestions for further work.

10.1 Conclusions

The physical model is more complex, with a reduction in pessimistic assumptions that could have significantly overestimated cable temperatures and degradation. The temperature modelling has been adapted through making targeted changes to standard formulae to be more representative of the cable-soil heat transfer problem. Considering transient temperatures has significantly reduced the maximum temperatures predicted for a given demand profile, and gives a more realistic value for the time a cable spends at these maximum temperatures. Joints are now explicitly modelled as hotter than their surrounding cables, linking the physical modelling clearly to ENW's observations of joints as the failure location in the vast majority of their LV ways.

Damage predictions in the physical model have benefited from detailed reviews of reports from actual failed cables and joints provided by other industry stakeholders. Chemical damage has been added to the mechanisms considered, and the model methods are now capable of considering both older PILC cables in addition to the XLPE cables which were the focus of the PoC work in LV Predict 1.

Simulations of cables and joints are now based on topologies of real networks in ENW's area, with simulated currents resulting from emulators trained on load flow results. This means the conversion from demand to current is much more realistic and the complete set of simulations reflects the variation in topologies (and therefore the variation in relationship between demands and currents) seen in reality. Also due to the inclusion of real network topologies in the simulation process, a range of joints with different numbers and types of cables connected have been modelled, with realistic numbers of customers connected and accurate geographic inputs (soil temperatures, latitudes and elevations). This is a significant improvement over the set of random individual joints modelled in LV Predict 1. To account for the variability in LV networks across ENW's area and capture a wide range of possible scenarios resulting from variations in LCT uptake and random demand behaviours, large-scale simulation was conducted with over 140,000 cable-years simulated. This is significantly larger in scale than the 24,000 cable-years simulated in LV Predict 1 and to our knowledge the only detailed simulation of joint degradation for such a large sample of networks.

The complete modelling framework is now comfortably beyond the PoC stage. Task 1.1 of the project has been achieved through major progress on both physical and statistical modelling methods. Task 1.2 has been achieved through applying the methodology steps using real ENW network topologies, demonstrating the model usage, and investigating sensitivities to key inputs.

Caveats and assumptions remain (listed in Sections 3.5 and 4.2), and it is recommended that the output fitted models from this work are not used in isolation from the full methodology (due to the risks and benefits discussed in Section 9.3). However, the analyses using the proposed methodology from LV Predict 2 would now be suitable to inform decision-making when implemented, subject to verification and calibration against observed fault rates. Furthermore, the research progress and findings allows ENW to plan both future research and asset management decisions.

10.2 Recommendations

The next steps for the LV Predict methodology could include further Research and Development (R&D), or software architecture work to look at packaging the methods into an asset management tool to enable continuous decision-making.

It is recommended that any future work be prioritised using the outcomes of the CBA methodology work — especially noting that implementing the full methodology, fitting from base data, is recommended. The CBA methodology report [1] investigates the level of accuracy in predictive targeting of proactive maintenance needed in order to deliver benefits over the current reactive approach. Therefore the specific R&D recommendations presented below should be reviewed against the CBA methodology findings, so that any future projects focus on topics that could give the greatest benefit in informing asset management planning. ENW's closedown report on this NIA project is expected to revisit these recommendations alongside the findings of the CBA work. The recommendations are:

- ▶ Failure validation data would provide greater confidence in the model damage predictions. At the highest level, this could be a focus on LV Predict's estimates of damage for recently faulted joints. As a predictor of long-term degradation, the LV Predict model could be used to triage faulted joints into one-off failures (due to e.g. poor initial manufacturing quality, not covered by LV Predict) and areas at risk of long-term issues due to high-demand or location-specific soil conditions (the factors driving LV Predict's models). Tracking the agreement of LV Predict with future joint failures, especially recurrent failures, would validate this triage process. A gap in modelling understanding remains between the long-term accumulation of damage and the triggering of failure, which could be delayed until, for example, a period of wet weather causing water ingress. Validation data that addresses this gap in understanding would be particularly beneficial.
- ▶ A more detailed validation of physical model specifics would involve further cable 'post-mortems' like those mentioned in the SYNAPS 2 NIA projects [2–5] to check for evidence of degradation. For robust validation, these investigations would need to be supported by data on the cable's history, including age and historical demand, in order to validate the physical model's links between temperature and damage. For such detailed validation to be viable, significantly more instrumentation may be needed on the LV network than currently exists.
- ▶ Improvements to specific areas of the physical model would be gained from testing more representative materials to reduce the uncertainty on material properties in individual degradation formulae.
- ▶ The simulations would benefit from improved input demand data. Recent datasets with many individual time series, a long time span and good data availability were difficult to obtain but would improve the representativeness of the results. In particular, more recent smart meter demands and a larger set of heat pump demands would be beneficial. Work to better understand the behaviour of individual heat pumps, their weather dependence and possible clustering into behavioural or housing types is recommended. The extension to include industrial demands would improve the accuracy of the input demand data, but would require insight into the demands associated with each industrial connection — possibly through classification into demand types.
- ▶ Further work to improve the discrimination of the damage model predictions would greatly improve the potential for the tool to provide informative outputs. We expect this improvement would most likely come from refinement of the temperature model, where further research could consider accounting for transient temperature effects more explicitly.

- ▶ Further investigation of the root causes of poor predictive performance for the 'complex' networks, including inclusion of network topology features based on network impedances. If this does not provide sufficient improvement, it may be recommended to simulate these network types in full rather than use the statistical models to make predictions.
- ▶ The combined physical-statistical model is necessary for these LV assets primarily due to the low level of monitoring data. Similar projects on other network assets (especially High Voltage (HV) equipment) have been able to adopt a fully statistical approach, avoiding the uncertainties and validation challenges of explicit physical modelling. A purely statistical approach remains an option for the LV network if the monitoring strategy for these assets were to change significantly in future regulatory periods. However, the combined physical-statistical methodology developed in this project provides the additional flexibility to model conditions not experienced on the network, allowing consideration of both historic and future possible scenarios.

11 References

1. Taylor M, Tawn R, and McFadzean G, LV Predict 2 — Cost Benefit Analysis Methodology Report, FNC 020424/60409R, Issue 1, February 2026.
2. Evans I, SSEN — Lyndhurst — SYNAPS 2; Cable and Joint Examinations: Examination of pre-breakdown joint, RINA Tech UK Limited Report 2022-0117 Rev. 1, February 2022.
3. Evans I, SSEN — Lyndhurst — SYNAPS 2; Cable and Joint Examinations: Examination of pre-breakdown LV cable Joints, RINA Tech UK Limited Report 2023-0005 Rev. 0, January 2023.
4. Ren X, and Stubbs J, UK Power Networks — Three Bridges — Laboratory Examination of LV Cable and Joint, RINA Tech UK Limited Report 2024-0761 Rev. 0, May 2024.
5. Energy Networks Association, Smarter Portal. https://smarter.energynetworks.org/projects/nia_ssen_0051/, accessed 03/12/2025.
6. ENW, NARM EJP 3 - LV Cable Replacement, 2021.
7. UK DNOs, DNO Common Network Asset Indices Methodology, Version 1.1, Ofgem, 2015.
8. Klerx M, Morren J, and Slootweg H, Investigation of failures of low voltage grid components. In *52nd International Universities Power Engineering Conference (UPEC)*, 2017.
9. Lee N, Taylor M, Finley J, Edwards G, Tawn R, McFadzean G, and Hay S, LV Predict — Final Report, FNC 012225-101/54232R, Issue 1, August 2023.
10. CIGRE, Guide for thermal rating calculations of overhead lines, 601, 2014, <https://www.e-cigre.org/publications/detail/601-guide-for-thermal-rating-calculations-of-overhead-lines.html>.
11. British Standards Institution, Electric cables. Calculation of the current rating. Part 2-1: Calculation of Thermal Resistance, BS IEC 60287-2-1: 2023.
12. British Standards Institution, Guidance concerning the permissible temperature rise for parts of electrical equipment, in particular for terminals, PD IEC/TR 60943:2009.
13. American National Standards Institute, ANSI/NEMA MG 1-2021: Motors and Generators, 2021.
14. British Geological Survey, Georeports: Temperature and Thermal Properties (Detailed). Nottingham.
15. Zhao Y, Si B, Zhang Z, Li M, He H, and Hill R L, A new thermal conductivity model for sandy and peat soils. *Agricultural and Forest Meteorology*, 274(1), 95–105, 2019.
16. Matmake, Specific heat capacity of soils. <https://matmake.com/properties/specific-heat-capacity-of-soils.html>.
17. IES, Thermal conductivity, specific heat capacity and density. https://help.iesve.com/ve2021/table_6_thermal_conductivity__specific_heat_capacity_and_density.htm.
18. Office M, Met office midas open: Uk land surface stations data (1853-current). <http://catalogue.ceda.ac.uk/uuid/dbd451271eb04662beade68da43546e1>.
19. Parker D, Legg T, and Folland C, A new daily central england temperature series, 1772-1991. *Int J Clim*, (12), 317–342, 1992.

20. Office for National Statistics, Rural urban classification (2021) of LSOAs in England and Wales. 2021, https://geoportal.statistics.gov.uk/datasets/9dbf7613cbb147b8bb8627ddb3568cff_0/explore, licensed under the Open Government Licence v.3.0. Contains OS data © Crown copyright and database right 2025.
21. Scottish Government, Urban rural classification 2022. 2022, <https://www.gov.scot/publications/scottish-government-urban-rural-classification-2022/>.
22. Electricity Specification 400R10: Cold-Pour Resin Compound for Enapsulating Joints on LV, 11kV, and 33kV Cables.
23. Harper C A, Handbook of Plastics, Elastomers, and Composites.
24. Betten J, Creep Behavior of Isotropic and Anisotropic Materials; Constitutive Equations, pp. 49–83. Springer Berlin Heidelberg, 2008, [dx.doi.org/10.1007/978-3-540-85051-9_4](https://doi.org/10.1007/978-3-540-85051-9_4).
25. Cui Y, Campbell J, Burley M, Patel M, Hunt K, and Clyne T, Effects of temperature and filler content on the creep behaviour of a polyurethane rubber. *Mechanics of Materials*, 148, 103461, 2020, [dx.doi.org/https://doi.org/10.1016/j.mechmat.2020.103461](https://doi.org/10.1016/j.mechmat.2020.103461).
26. PyPI, Rainflow. <https://pypi.org/project/rainflow/>.
27. for Testing A S, and (ASTM) M, ASTM E1049-85: Standard Practices for Cycle Counting in Fatigue Analysis, 2017.
28. Tan W, Na J, Wang G, Xu Q, Shen H, and Mu W, The effects of service temperature on the fatigue behavior of a polyurethane adhesive joint. *International Journal of Adhesion and Adhesives*, 107, 102819, 2021, [dx.doi.org/https://doi.org/10.1016/j.ijadhadh.2021.102819](https://doi.org/10.1016/j.ijadhadh.2021.102819).
29. Lv H, Lu T, Xiong L, Zheng X, Huang Y, Ying M, Cai J, and Li Z, Assessment of thermally aged XLPE insulation material under extreme operating temperatures. *Polymer Testing*, 88, 2020.
30. Kim J, Kim W, Park H S, and Kang J W, Lifetime assessment for oil-paper insulation using thermal and electrical multiple degradation. *Journal of Electrical Engineering and Technology*, 12, 840–845, 2017, [dx.doi.org/10.5370/JEET.2017.12.2.840](https://doi.org/10.5370/JEET.2017.12.2.840).
31. Zapf M, Blenk T, Müller A C, Pengg H, Mladenovic I, and Weindl C, Lifetime Assessment of PILC Cables with Regard to Thermal Aging Based on a Medium Voltage Distribution Network Benchmark and Representative Load Scenarios in the Course of the Expansion of Distributed Energy Resources. *Energies*, 14(2), 2021.
32. Summers R, Standard Technique: SD8B/4 (Part 1): Relative to LV Underground Cable Ratings, SD8B/4, July 2016.
33. Electricity North West, 1 to 4 way Single Phase CNE Service Breech Joint to 3 Core Waveform Cable, Jointing Instruction S-M/1P/BJ/201A SM1.
34. Electricity North West, 3 Core Waveform to 4 Core PILC Straight Joint, Jointing Instruction M/3P/SJ/204A M4.
35. Thorne and Derrick International, www.powerandcables.com, Sicame PCM11-12-13 Insulation-Piercing Aluminium LV Main Straight Connector.
36. Aluminium 6061-T6; 6061-T651 Material Data Sheet. <https://asm.matweb.com/search/SpecificMaterial.asp?bassnum=ma6061t6>.

37. Summers P, Chen Y, and C R, Overview of aluminum alloy mechanical properties during and after fires. *Fire Sci Rev*, 4, 3, 2015, [dx.doi.org/https://doi.org/10.1186/s40038-015-0007-5](https://doi.org/10.1186/s40038-015-0007-5).
38. Shivakumar S, Sharan A, and Sadashivappa K, Constitutive modeling of creep properties of Aluminum 6061 Alloy. *IOP Conference Series: Materials Science and Engineering*, 310(1), 012027, 2018.
39. Yahr G, Fatigue design curves for 6061-T6 aluminum. *Journal of Pressure Vessel Technology*, 119(2), p. 211–215., 1997.
40. EPRI, Environmental Impacts of Lead from Paper-Insulated Lead-Covered Cable, Technical Update Report 1009513, April 2004.
41. Northern Powergrid, CLNR-G026: Customer-Led Network Revolution - Project closedown report, 2015.
42. UK Government Department for Energy Security and Net Zero, DUKES historical electricity data: 1920 to 2023. 2024, <https://www.gov.uk/government/statistical-data-sets/historical-electricity-data>.
43. Western Power Distribution, Electric Nation: Powered up summary report, 2019.
44. Electric vehicle energy tariffs. <https://www.moneysavingexpert.com/utilities/ev-energy-tariffs/>.
45. Energy Systems Catapult, Electrification of Heat Demonstration Project summary report, 2024.
46. TNEI for National Grid ESO, QWID FLEXER final report, 2024.
47. Wheatcroft E, Dent C, and Wilson A, Rescaling of historic electricity demand series for forward-looking risk calculations. In *2022 17th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pp. 1–6, 2022, [dx.doi.org/10.1109/PMAPS53380.2022.9810599](https://doi.org/10.1109/PMAPS53380.2022.9810599).
48. Lamberti F, Dong C, Calderaro V, and Ochoa L F, Estimating the load response to voltage changes at uk primary substations. In *IEEE PES ISGT Europe 2013*, pp. 1–5, 2013, [dx.doi.org/10.1109/ISGTEurope.2013.6695466](https://doi.org/10.1109/ISGTEurope.2013.6695466).
49. Powergrid N, Code of Practice for Underground Cable Rating and Parameters. 2021.
50. Danka T, and Horvath P, modAL: A modular active learning framework for Python. <https://github.com/cosmic-cortex/modAL>, available on arXiv at <https://arxiv.org/abs/1805.00979>.
51. Shi Y, Ke G, Soukhavong D, Lamb J, Meng Q, Finley T, Wang T, Chen W, *et al.*, lightgbm: Light Gradient Boosting Machine. 2025, <https://github.com/Microsoft/LightGBM>.
52. Bloodgood M, and Vijay-Shanker K, A method for stopping active learning based on stabilizing predictions and the need for user-adjustable stopping. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL-2009)*, 2009, <https://arxiv.org/abs/1409.5165>.
53. ENW, DFES 2024 LV workbook. 2024, <https://www.enwl.co.uk/get-connected/network-information/dfes/>.
54. Wood S, Generalized Additive Models: An Introduction with R. Second edition, Chapman and Hall/CRC, 2017.
55. Stogiannopoulos T, Cheimariotis G A, and Mitianoudis N, A study of machine learning regression techniques for non-contact spo2 estimation from infrared motion-magnified facial video. *Information*, 14, 301, 2023, [dx.doi.org/10.3390/info14060301](https://doi.org/10.3390/info14060301).

A Physical model demonstrator code structure

The formulae and algorithms required for the physical model calculations were implemented in a demonstrator Python application. This annex documents the architecture and implementation details, serving as a point of reference for future developers looking to utilise the LV Predict 2 methodology.

A.1 Code structure

The demonstrator code has been designed with an Object-Oriented Programming (OOP) structure. This modular approach implements three classes: Joint, Cable and Cable Specification. The Joint class acts at the centre of the modelling and is where damage is calculated. The structural relationship between these elements is illustrated in Figure 105. The core architecture of the full demonstrator code is described in a Unified Modelling Language (UML) diagram in Section A.3.

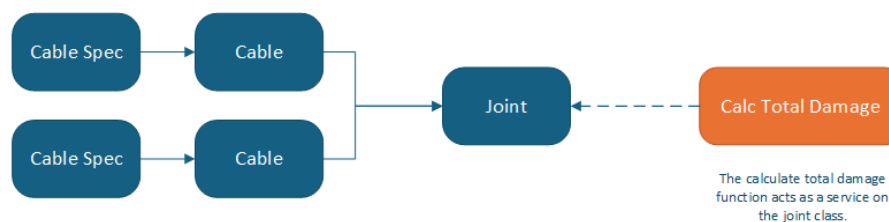


Figure 105: Code structure diagram.

A.1.1 Cable specification class

The cable specification class stores cable material properties and implements methods for calculating any derived material properties and temperature dependent quantities. This can be subclassed to customise for different conductor materials. Two were included for this work:

- ▶ Copper Conductor
- ▶ Aluminium Conductor

A.1.2 Cable class

The cable class allows for each cable in a joint to be modelled by a distinct object. The class is responsible for:

- ▶ Storing time series of demand and temperature.
- ▶ Calculating steady state and transient temperatures.

Temperature calculations are based upon ohmic heating and ground temperatures. Steady state temperatures are calculated first using demand, which is quantified using a time series of currents. This is then used as a basis for calculating transient temperature response by use of the thermal time constant as defined in Chapter 3.

A.1.3 Joint class

A joint object is composed of an arbitrary number of cables. In the simplest scenario this would be just two - one upstream and one downstream. However, the code allows for more complex arrangements to be modelled (e.g. two upstream, one downstream) if required. The joint temperature is taken as the maximum temperature over all of the joint's cables at each time point is taken. This value is used for calculating stresses and damage, as explained in Chapter 3. The calculation of additional heating due to contact resistance in joint connector bolts is also a method of the joint class.

A.2 Damage calculation

As discussed in Chapter 3, the joint class implements a method for calculating the total damage accumulated from different damage sources. Starting with a time series of demand on each cable in the joint, different damage metrics are calculated. The full data flow through the application is illustrated in Figure 106, which breaks down into the following series of operations within the code:

- ▶ Thermal stresses are calculated using temperature aggregated over the time series of all cables in the joint.
- ▶ Rain flow counting is used to extract stress cycles from the resultant time series.
- ▶ Dwell levels and times extracted.
- ▶ The total damage function is then used to calculate damage using the thermal stresses and dwell times as a basis.

The function returns a data structure containing aggregated quantities as well as the individual damage components presented in Section 3.3.

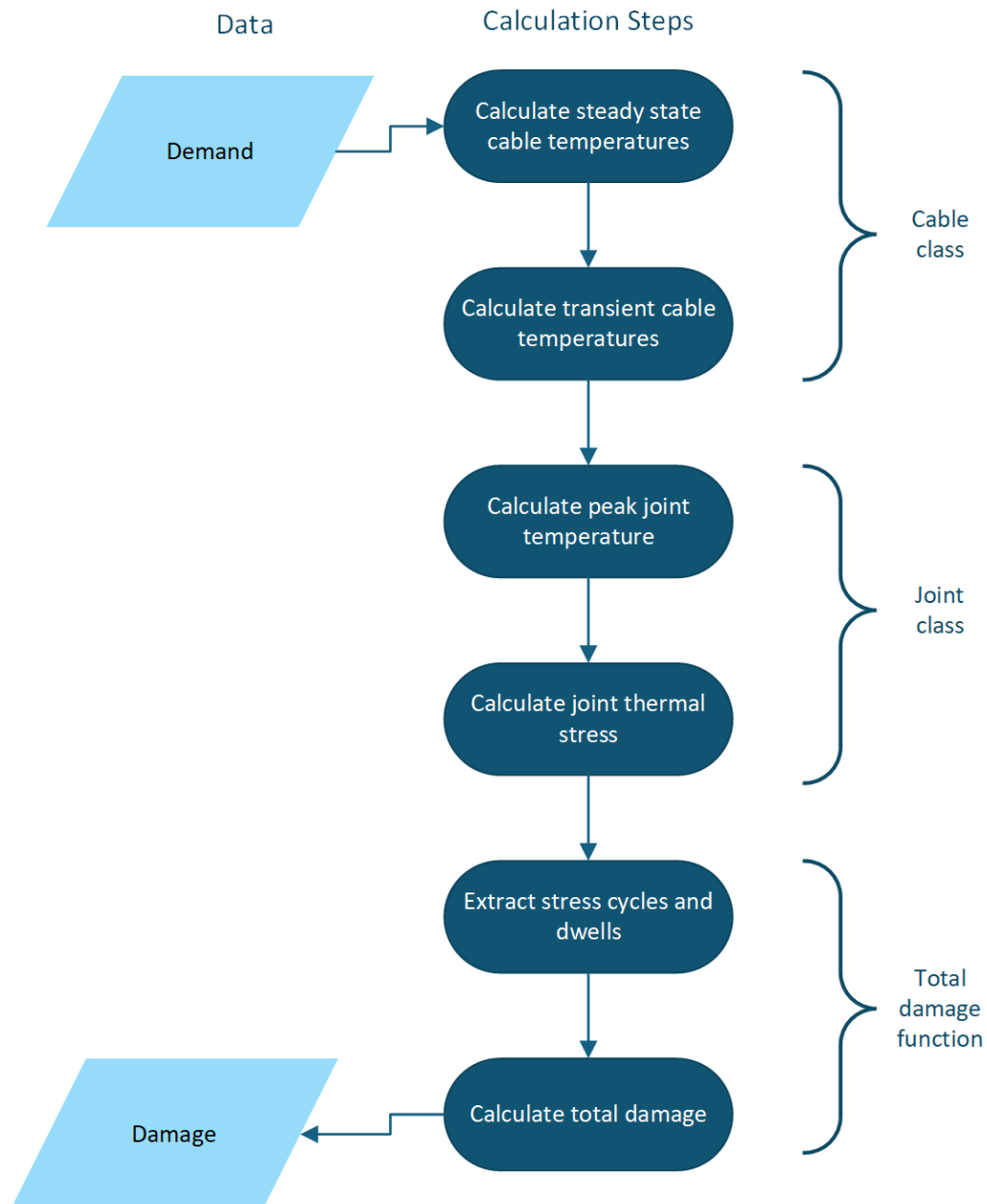


Figure 106: Data flow diagram.

A.3 Code structure UML

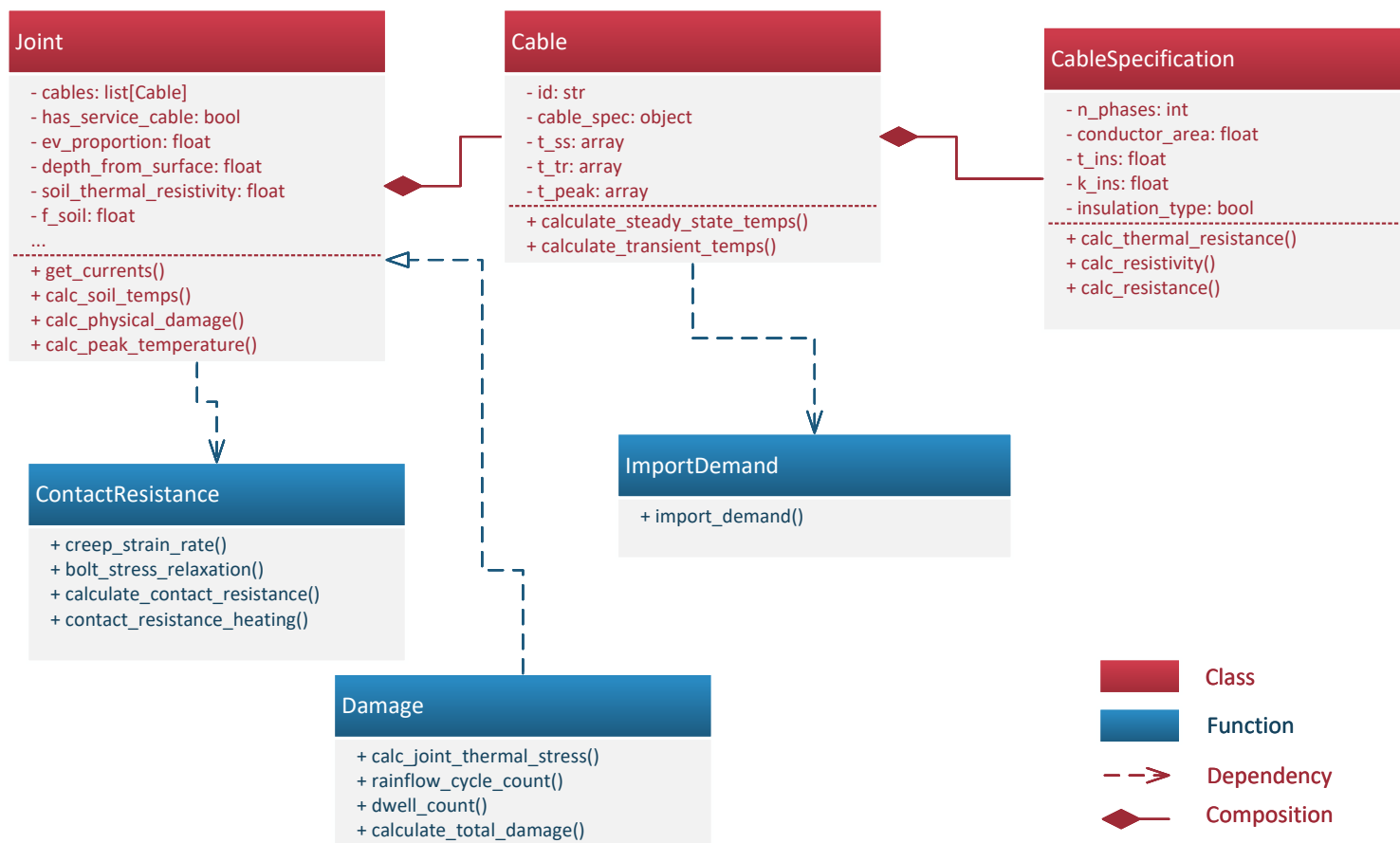


Figure 107: UML diagram describing the classes and functions that make up the demonstrator code. Relationships between different components are also shown.

B Detail on technical terms

B.1 GAM

A Generalised Additive Model (GAM) is a type of flexible statistical model that extends traditional linear regression to allow smoothly varying, nonlinear relationships between the input features and the target variable. Instead of fitting a single coefficient for each feature and forcing the relationship to follow a straight line as linear regression does, GAMs use a *semi-parametric* approach. They fit coefficients to a set of basis functions, and by controlling the relative size of each basis function a wide range of forms of relationship between inputs and the target can be modelled. This is demonstrated in figure 108. Changing the number of basis functions controls the allowed ‘wiggleness’ of the fitted relationship, with a preference for a low number of basis functions (typically 4–8) to prevent overfitting. Splines can be constrained to be cyclical (e.g. so for a calendar feature the value for the 31st December would align with the value for the 1st January). Interactions between input variables can also be included through two-dimensional splines. Thus, domain knowledge can be used to specify the form of suitable splines, but with the actual shape of the relationship dependent on the data. The marginal fitted spline relationships can also be extracted and plotted, maintaining interpretability of the model. They are implemented in R with a range of available basis functions, link functions (transformations applied to the target function) and fitting methods through the *mgcv* package [54].

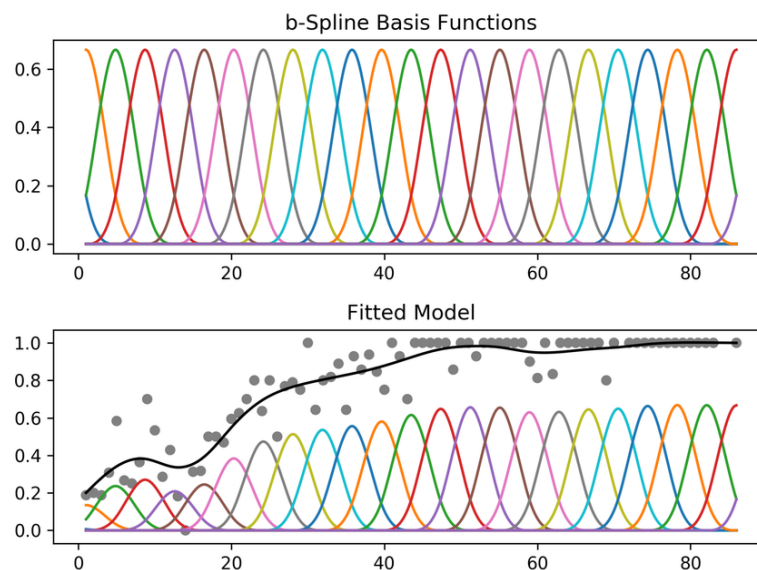


Figure 108: Above: Set of basis functions. Below: Basis functions fitted to data. When the basis functions are summed, they produce overall relationship (black line). Plot reproduced from [55].

In terms of notation, $s(x)$ is used to represent a one-dimensional spline fitted to the parameter x , and $ti(x, y)$ denotes a two-dimensional spline of the interaction between variables x and y . In the *mgcv* package, there is also a full tensor product smooth, $te(\cdot)$, which includes both the interaction term and the individual effects (i.e. $te(x, y) = ti(x, y) + ti(x) + ti(y)$).

B.2 Elastic net regularisation

In general, statistical models are fitted by estimating the parameters associated with the input features. To optimise the model parameters for the best fit, an objective function is minimised that is usually related to the agreement between the model predictions and the target values. For example, standard linear regression minimises the square of the residuals:

$$S = ||y - X\beta||^2 \quad (41)$$

where $X\beta$ are the model predictions y are the target values.

In some cases, there may be benefit to imposing additional penalties within this objective function. In particular, if many features are supplied where some features are unlikely to be informative, a penalty on the absolute value of the coefficients β will allow the coefficients of the least informative features to be shrunk to zero (or very close to zero), effectively automatically removing non informative features from the model and providing a better fit. A penalty on the square of these coefficients improves the shrinkage effect when multiple features are highly correlated. Overall, this penalty is known as elastic net regression.

$$S = ||y - X\beta||^2 + \lambda_1 ||\beta|| + \lambda_2 ||\beta||^2 \quad (42)$$

The coefficients λ_1 and λ_2 control the relative size of the penalty terms, effectively determining the strength of the regularisation (i.e. how many terms are forced to zero). These hyperparameters can be tuned through cross-validation to produce the best fit for the particular dataset and application.

B.3 Cross-validation

Cross-validation is a technique used to ensure models are evaluated out-of-sample so reported performance is representative of real world performance. Its structure allows out-of-sample predictions to be made for all available data points, maximising the use of the data for both model training and testing. The data is split equally into k folds, with each data point assigned to one fold. Models are trained on all but one fold, and then predictions made for the fold that was left out. This process is repeated for each fold, giving k fitted models and out-of-sample predictions for every time point. This process is illustrated in figure 109.

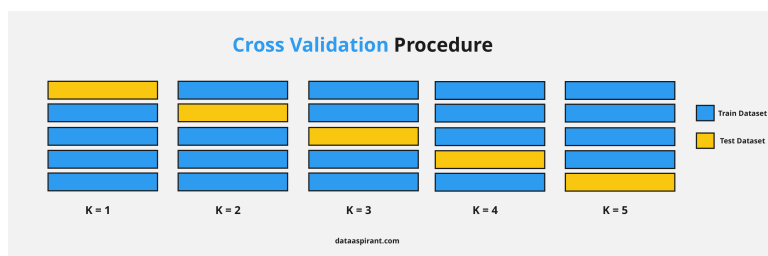


Figure 109: Diagram showing training and testing combinations across all folds. Plot reproduced from <https://dataaspirant.com/cross-validation/>.

C Table of soil properties used in modelling

Soil type	Dry thermal resistivity (°C m/W)	Wet thermal resistivity (°C m/W)	Dry specific heat capacity (J/kg°C)	Wet specific heat capacity (J/kg°C)	Dry density (kg/m ³)	Wet density (kg/m ³)
PEAT	12.50	2.22	3300	3650	700	1100
VARIED, LOCALLY PEATY	1.58	0.95	2684	2999	845	1286
SAND	1.30	0.40	1190	1480	1800	2000
SAND TO SANDY LOAM	1.24	0.48	1102	1372	1670	1961
SAND TO LOAM	1.19	0.59	1014	1263	1540	1921
LOAM TO SANDY LOAM	1.14	0.76	925	1155	1410	1882
LOAM	1.10	1.10	837	1046	1280	1842
CLAYEY LOAM TO SANDY LOAM	1.08	0.67	1092	1359	1627	1947
LOAM TO CLAYEY LOAM	1.04	0.91	940	1172	1410	1882
CLAY TO SANDY LOAM	1.03	0.59	1132	1407	1670	1961
LOAM TO CLAY	0.99	0.78	1044	1298	1540	1921
CLAY TO CLAYEY LOAM	0.94	0.68	1147	1424	1670	1961
SILTY LOAM TO SANDY LOAM	0.94	0.67	1344	1521	1523	1854
LOAM TO SILTY LOAM	0.91	0.91	1256	1413	1393	1814
CLAYEY LOAM TO SILTY LOAM	0.87	0.78	1359	1539	1523	1854
SILTY LOAM	0.78	0.78	1675	1779	1505	1786
CLAY TO SILT	0.72	0.60	1881	2031	1765	1865
SILT TO SILTY LOAM	0.68	0.68	2093	2146	1618	1758
CLAY	0.90	0.60	1250	1550	1800	2000
SILT	0.60	0.60	2512	2512	1730	1730

Table 12: Values for soil properties used in thermal modelling of cables and joints.



Narrow Quay House
Bristol
BS1 4BA

Tel: +44 (0)117 922 6242

fnc.co.uk

Offices at:

Bristol, Burton-on-Trent, Cambridge, Culham, Glasgow,
Gloucester, Leatherhead, Lincoln, Manchester, Middlesbrough,
Plymouth and Portsmouth