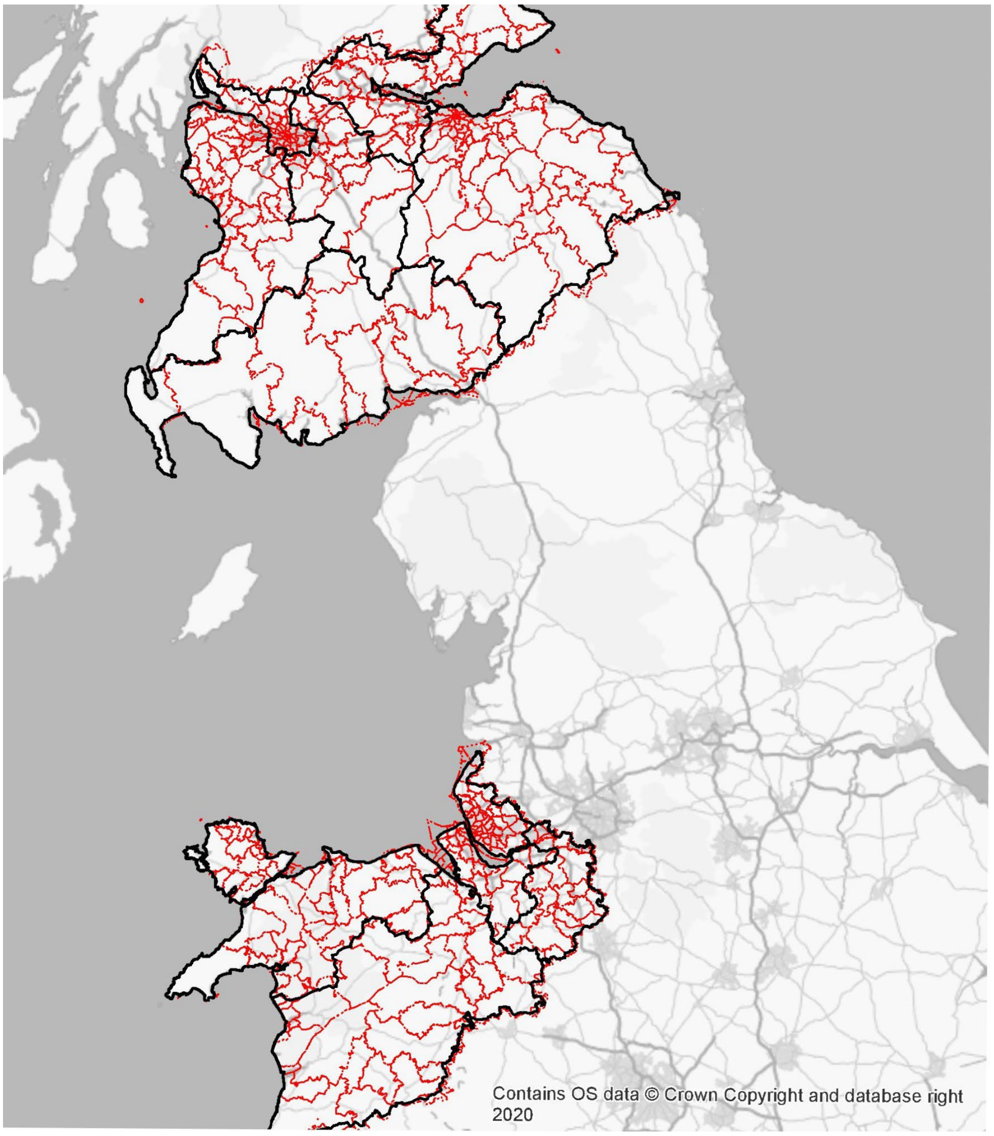
Scottish Power Energy Networks

Predict4Resilience

Historical Network Fault Data Review Report

Reference:

ISSUE 1 | 21 April 2022



This report takes into account the particular instructions and requirements of our client. It is not intended for and should not be relied upon by any third party and no responsibility is undertaken to any third party.

Job number 286969

Ove Arup & Partners Limited
6th Floor 3 Piccadilly Place
Manchester
M1 3BN
United Kingdom
arup.com

Contents

Tables

Figures

Executive Summary

Predict4Resilience is an innovation project sponsored by the Strategic Innovation Fund (SIF) under the category of Data and Digitalisation, which aims to benefit from the recent advances in supercomputing and numerical weather prediction to combine state-of-art weather forecasting and novel statistical post-processing to predict the impacts of weather conditions on distribution networks within a window of 7 to 14 days, in the form of the expected number of network faults within a geographical region (e.g., a district). This will assist managers and engineers to make more informed decisions leading to better preparedness against adverse weather events and increase in the efficiency of operation practice; and ultimately deliver benefits to the customers (e.g., better management of planned maintenance, reducing customer minutes lost, etc.).

Historical network fault data play a fundamental role in the data-driven approach that the Predict4Resilience project is taken. Thus, the availability and quality of the data determines the feasibility and effectiveness of transferring weather forecasts into fault forecasts. With the objective of predicting the number of faults having main causes related to weather conditions in a geographical region in anticipation of adverse weather events, the forecast model needs to learn and build the statistical association between weather variables (e.g., wind speed, air surface temperature, precipitation, etc.) and the number of network faults in each region based on historical weather and network fault data. This statistical association of weather and network fault will be at the core of the forecast model driving the quality of the fault forecast.

Therefore, in the Preparation Phase of the project, a detailed review of the historical network fault data has been performed. The specific data required, the data's purposes, expected data quality, and the quality of the available data are summarised as follows. On the other hand, the minimum requirement for the data and possible mitigation measures are discussed in detail in Section 2.2.

Data Component	Purpose	What Good Looks Like	Is the Available Data Sufficient?
Timestamp for each network fault	An accurate timestamp representing when the network fault occurred to build the temporal link between weather variables and network faults.	A timestamp with a time and a date, providing the flexibility for grouping faults as needed by the forecast model.	Yes.
Record of the main cause of each network fault	An accurate description of the main cause of the network fault to enable the identification of the core weather variable(s) associated with the network fault, and to build the link between weather variable(s) and network faults.	A detailed description of the main cause, covering both the primary weather variable(s) at the time of the network fault and the secondary or more direct cause of the network fault.	Yes.
Record of the geographical location of each network fault	An accurate identification of the geographical location of the network fault to enable the geographical link between weather variable(s) and network faults, and to build the between weather variable(s) and network faults.	A pair of coordinates (e.g., latitudes and longitudes, grid references), or post codes, providing the flexibility for grouping faults as needed by the forecast model.	Yes, at district level, but this will not take the full advantage of the weather data. Finer granularity is preferred.
Record of the asset involved in each network fault	An accurate record of the asset involved in the network fault to support the identification and verification of the main cause of the network fault, and to build the link between weather variable(s) and typical types of assets that can be affected.	A description of the asset involved in the network fault, e.g., the asset type, the asset's installation environment, etc.	Yes, for typical asset type. Detailed exploration will be needed in line with the need of the forecast model.
Record of the impact of each network fault	A quantitative representation of the magnitude of the network fault's impact on the affected network asset and customers, respectively, to enable the estimate of potential benefits that may be brought by the fault forecast.	– Network asset: scale and nature of the restoration – Duration of the network fault before being restored to normal condition – Number of customers affected	Yes, the customer hours lost, and customer interruption are available for each network fault.

As noted in the table above, the detailed data review finds that the available historical network fault data are of relatively good quality, and timestamps and geographical locations (though at district level) are generally recorded in a consistent manner. Nonetheless, two thirds of the Grid References X and Y are blanket, whilst some of them may contain errors.

Furthermore, it is revealed that the top “Main Cause” for network faults is “Deterioration due to Ageing or Wear (excluding corrosion)”, whilst “Cause Unknown” has the second and third highest number of recorded network faults for SPD and SPM, respectively. In the meantime, “Wind and Gale (excluding Windborne Material)” is the top one weather-related “Main Cause” for network faults, which is followed by “Lightning” and “Snow, Sleet and Blizzard”. Geographically speaking, the top three SPD districts with high number of weather-related network faults include Dumfries, Ayrshire and Clyde South, and Central and Fife; similarly, the top three SPM districts with high number of weather-related network faults include North Wales, Dee Valley and Oswestry, Mid Wales. It is also found that three quarters of the top three weather-related network faults would have comparable magnitude of impact on the customers from Customer Interruption and Customer Minutes Lost’s point of view.

For the Discovery Phase, the historical network fault data have been grouped by their “Main Cause”, and subsequently the number of faults has been counted on an hourly basis in each district. These numbers of faults are used to establish the statistical association with the relevant weather data, in order to provide a proof of concept for the Discovery Phase. Meanwhile, the Customer Hours Lost has been used to estimate the potential benefits that could be brought by the fault forecast.

For the next phase of the project, there will be a need to explore how the granularity of the geographical locations of the network faults can be improved in order to take the advantage of the weather data and better suit the need of the forecast model. Similarly, depending on the development of the forecast model, it may be worth understanding how information of the assets involved in the network fault can be used to enhance the statistical association between weather variables and network faults. Moreover, the next phase will require an automated data processing to prepare the network fault data for the forecast model.

In summary, the historical network fault data is of relatively good quality, and there are clear opportunities for improvement to suit the emerging need and requirement of the forecast model including the incorporation of other supplementary data (e.g., land cover, vegetation survey, satellite imagery, etc.), as the forecast model is properly investigated and developed in the next phase of the project.

1. Introduction

The project of Predict4Resilience is UKRI funded collaboration among Scottish Power Energy Networks (SPEN), University of Glasgow (UOG), Met Office and Arup to develop a "Weather-Fault Tool", which aims to use the recent advances in supercomputing and numerical weather prediction to combine state-of-the-art weather forecasting and novel statistical post-processing to predict impact of short-term weather conditions on the distribution network (i.e., number of weather-related faults per SPEN district) in a window of 7 to 14 days. This project aims to empower engineers in SPEN's Control Room and Asset Management team to improve their ability to improve the efficiency of operational practices, protect network assets, and ultimately reduce customer minutes lost, extend asset lifespan, reducing thereby both the associated cost and carbon footprint.

This "Weather-Fault Tool" will be able to improve the preparedness of SPEN's Control Room against the faults associated with adverse, severe, and extreme weather events, and the focus will be on forecasts 7 to 14 days in advance. This can improve SPEN's use of available data as well as broaden the spectrum of data that may be made available by third parties such as the Met Office. These techniques that will be implemented in the "Weather-Fault Tool" will convert weather forecasts into impact forecasts, in this case forecasts of the number of network faults, which will provide probabilistic forecasts of the number of faults resulting in better decision-making than using raw weather forecasts only (current BAU). As a result, engineers will be able to make informed decisions based on actionable fault and risk predictions communicated to the control room.

In line with the resilience framework proposed by the National Infrastructure Commission, this "Weather-Fault Tool" will be able to improve anticipation of weather-related shocks and stresses and expedite recovery attributed to improved planning and mobilisation of resources. Further, this is envisaged to support lessons learnt sessions post-event and may contribute to the adaptation and transformation of both planning and operation of the distribution network.

The Discovery Phase will take existing network data sets, coupled with weather data acquired by UOG and assess if this data is sufficient to support the project aims, as well as identify the gap to realise the required format and volume.

We, Arup, understand that it provides both technical and programme benefits by performing a data review prior to the commencement of the Predict4Resilience Discovery Phase. Therefore, a Preparation Phase has been proposed to cover the scope as follows.

- Task 1: Data Collection and Examination
- Task 2: Data Alignment
- Task 3: High-level Due Diligence for Software and Modelling Techniques
- Task 4: Data Review Report
- Task 5: Mobilisation for the Discovery Phase

1.1 Scope of Services

The scope of services included for the Preparation Phase is detailed as follows.

1.1.1 Task 1: Data Collection and Examination

We will collect the data made available by SPEN, which are intended for use in the Discovery Phase and will examine the quality of the collected data focusing on their nature, contents, consistency, and format, among others. We will identify any bad data points and gaps and recommend the suitable data to be used in the Discovery Phase.

1.1.2 Task 2: Data Alignment

We will review the geographical and chronological alignment of the collected data in order to identify any key issues associated with establishing correlation among weather conditions, network conditions (e.g., faults or outages, etc). We will review spatial and temporal identification of the collected data and agree with SPEN and the project team on the parameters for grouping the collected data. We will prepare the collected data in order to allow the UOG project team to process the data against the weather data that they will have collected.

1.1.3 Task 3: High-level Due Diligence for Software and Modelling Techniques

We will, in collaboration with the project team, review the software and modelling techniques that will be appropriate for the investigation during the Discovery Phase. We will investigate the required interface between the weather forecast and any required network-related study.

1.1.4 Task 4: Date Review Report

We will prepare a Data Review report to document and summarise the findings from Tasks 1 to 3, including the status of the collected data, the suitability of the collected data, and the merits of different software and modelling techniques in the context of the Predict4Resilience project.

1.1.5 Task 5: Mobilisation for the Discovery Phase

We will identify and mobilise the relevant project delivery team within Arup for the Discovery Phase, based on the findings of this data review and the engagement with SPEN and the project team during the Preparation Phase.

1.2 Report Summary

This report is the deliverable for the Preparation Phase under Task 4. The report aims to provide both an overview and detailed analysis of the data provided by SPEN, to document the discussions on software and modelling techniques, and a short summary of the findings.

2. Historical Network Fault Data

This section firstly outlines the purpose of the historical network fault data, the expected data quality and defines the corresponding minimum requirement to meet the objective of the Predict4Resilience project. Afterwards, a general overview of the historical network fault data provided by SPEN is provided and the quality of the available data is detailed. Finally, a detailed analysis of the historical network fault data is presented to provide a comprehensive and in-depth understanding of the available data in both jurisdictions of SPD and SPM respectively, including main causes for faults and the recorded relationship between weather conditions and faults. The analysis examines the available data based on their seasonal characteristics, geographical distribution, and impact on customers, respectively.

2.1 Overall Purpose of Historical Network Fault Data

Historical network fault data are fundamental to the data-driven approach taken in the Predict4Resilience project. The availability and quality of the data determines the feasibility of transferring weather forecasts into fault forecasts.

More specifically, the objective of the forecast model is to predict the number of faults having main causes related to weather conditions in a geographical region in anticipation of adverse weather events. This means that the forecast model needs to learn and build the statistical association between weather variables (e.g., wind speed, air surface temperature, precipitation, etc.) and the number of network faults in each region based on historical weather and network fault data. This statistical association of weather and network fault will be at the core of the forecast model driving the quality of the fault forecast.

Additionally, it is worth mentioning that other data such as land cover, vegetation survey, asset type and condition, among others, may provide extra information to build links between the weather variables and the number of network faults. The extra information can better define the causality of network faults (e.g., high wind gust in presence of vegetation next to overhead asset could lead to a network fault, whereas high wind gust may or may not cause a network fault), and thus strengthen the statistical association of weather and network fault leading to improvement in the fault forecast.

2.2 Expected Quality and Minimum Requirement for Historical Network Fault Data

In addition to the overall purpose of the historical network fault data outlined in Section 2.1, the detailed components of the historical network fault data, their purposes, expected quality and their minimum requirement can be summarised in Table 1 as follows.

Table 1 Expected Quality and Minimum Requirement for Historical Network Fault Data

Data Component	Purpose	What Good Looks Like	Minimum Requirement
Timestamp for each network fault	This should be an accurate timestamp representing when the network fault occurred. This enables the temporal link between weather variables and network faults.	This can be a timestamp including both a time and a date, which provides the flexibility for grouping the faults according to different time windows as required by the forecast model.	This depends on the granularity required by the forecast model, e.g., hourly, 12-hour window, or daily, etc. As the aim is to provide a forecast for the number of faults during a specific day, the minimum requirement for the timestamp is a date when the network fault occurred.
Record of the main cause of each network fault	This should be an accurate description of the main cause of the network fault. This enables the identification of the core weather variable(s) associated with the network fault. Additionally, this can support incorporation of extra information (such as land cover and vegetation) to enhance the statistical association between weather variable(s) and network faults.	This can be a detailed description of the main cause, covering both the primary weather variable(s) at the time of the network fault and the secondary or more direct cause of the network fault. For example, the primary main cause could be wind gust, whereas the direct (secondary) cause would be a nearby object blown by the wind gust and fell on overhead line.	This should at least allow a clear identification of what the primary weather variable(s) is (or are) associated with the network fault. Some additional requirements can vary in accordance with how the forecast model captures extra information (e.g., secondary causes) to enhance the fault forecast.
Record of the geographical location of each network fault	This should be an accurate identification of the geographical location of the network fault. This enables the geographical link between weather variables and network faults. Additionally, this can support incorporation of extra information (such as land cover and vegetation) to enhance the statistical association between weather variable(s) and network faults.	This can be a pair of coordinates such as latitudes and longitudes, or grid references, or post codes, which provide high flexibility for grouping the faults according to different spatial granularity required by the forecast model.	This is determined by the spatial granularity considered by the forecast model. As the fault forecast is at least expected at SPEN's district level, the minimum requirement for geographical location of the network fault is the corresponding district.
Record of the asset involved in each network fault	This should be an accurate record of the asset involved in the network fault. This can support the identification and verification of the main cause of the network fault. This also enables the link between weather variables and typical types of assets that are or can be affected. Similarly, this can support incorporation of extra information (e.g., land cover and vegetation) to enhance the statistical association between weather variable(s) and network faults.	This can be a description of the asset involved in the network fault with details such as the asset identification in an asset management system, the asset type, the operational conditions when the network fault occurred, and the asset's installation environment, etc.	Depending on the forecast model's preference, the minimum requirement for asset information should be an identification of the asset type (e.g., overhead lines, outdoor switchgear/busbar, etc.).
Record of the impact of each network fault	This should be a quantitative representation of the magnitude of the network fault's impact on the affected network asset and customers, respectively. This enables the estimate of potential benefits that may be brought by the fault forecast. Additionally, this could enable statistical analysis of the weather variables' impact on the customers.	This includes the restoration to the affected network asset(s), the duration of the network fault before being restored to normal condition, the number of customers affected; This could also be the Customer Minutes Lost and Customer Interruption for each network fault.	The requirement for impact of the network faults may depend on what and how the benefits are estimated. As the aim is to at least capture the benefits from the customer services' point of view, the minimum requirement is to have the Customer Minutes Lost of the network fault.

2.3 Overview of Historical Network Fault Data

SPEN provided historical network fault data recorded by their Control Room for the jurisdiction of both Scottish Power Distribution PLC (SPD) and Scottish Power Manweb PLC (SPM) throughout Central and Southern Scotland, North Wales, Merseyside, Cheshire and North Shropshire. The historical network fault data are summarised in Table 2 for both SPD and SPM.

Table 2 Summary of SPD and SPM's Historical Network Fault Data

Network	Period	Total Number of Faults	Total Number of Categories for Main Cause
SPD	1 April 2010 – 1 April 2021	36,551	66
SPM	1 April 2010 – 1 April 2021	32,067	65

2.3.1 Data Content and Format

The historical network fault data are provided in two separate Excel spreadsheets for SPD and SPM, respectively. The data presented in the spreadsheets for SPD and SPM are organised in the same format, which cover the following headings.

- District
- Master Job Ref
- Date/Time of Incident
- Reporting Year
- Circuit Code
- HV Circuit Name
- Circuit Name (Incd.)
- Scheme Code
- Voltage
- Location
- Grid Reference X
- Grid Reference Y
- Main Cause
- HV Component
- Longest Stage Duration (hrs)
- CI (Incd)
- CHL(Incd)

The data recoded under the headings of "District", "Date/Time of Incident", "Main Cause", "Longest Stage Duration (hrs)", "CI (Incd)", and "CHL (Incd)" are important for the investigation during the Discovery Phase.

More specifically, SPEN districts recorded under "District" would be used as the geographical identification to build the association with historical weather data (e.g., areas where weather data will be extracted). On the other hand, timestamps recorded under "Date/Time of Incident" would be used as the basis for aligning chronologically the fault data with the historical weather data.

2.3.2 Data Consistency

The majority of historical network fault data for both SPD and SPM are recorded consistently under the aforementioned headings across different SPEN districts and reporting years, apart from "Grid Reference X" and "Grid Reference Y". The main inconsistency regarding "Grid Reference X" and "Grid Reference Y" is missing data, as shown in Table 3; additionally, other inconsistencies such as wrong or text entries for X and/or Y references are also found within the provided data.

Nonetheless, these missing and inconsistent entries for "Grid Reference X" and "Grid Reference Y" would not affect the investigation during the Discovery Phase. This is because it is deemed sufficient to analyse the faults at district level. It is important to highlight that the entries under "District" are assumed to be correct. Should there be any issue arising from the Discovery Phase specific investigation may be required to verify the underlying entries.

On the other hand, the discussions with SPEN's Geographical Information System (GIS) engineer suggest that there may be options to fill the gaps by using "Circuit Code", "HV Circuit Name", or "Location" information included in the data. This is deemed a time-consuming task which may not bring additional benefit for the Discovery Phase; however, these options may be further considered or re-evaluated should any need be identified in the next phase of the project.

Table 3 Fault Records with Blank Grid Reference X or Y

Network	Total Number of Faults	Number of Faults	Percentage
SPD	36,551	25,940	70.97%
SPM	32,067	24,031	74.94%

Moreover, it is found that there may be inconsistent entries under "Reporting Year", compared with entries under "Date/Time of Incident", as listed in Table 4 and Table 5 for SPD and SPM, respectively. There are three and thirteen inconsistent entries discovered in the historical network fault data for SPD and SPM, respectively.

Nonetheless, as mentioned previously, the timestamps recorded under "Date/Time of Incident" would be used to establish chronological correlation with historical weather data. As the timestamp under "Date/Time of Incident" is inherently more accurate, it is assumed that the entry under "Date/Time of Incident" would take precedence over the entry under "Reporting Year" in the presence of any discrepancy. It is also assumed that the entries under "Date/Time of Incident" are correct, and verification of these entries is not performed as part of the data review. Should the discrepancy identified below be worth further investigation specific analysis will be carried out in the next phase of the project.

Table 4 Inconsistent Entries in "Date/Time of Incident" and "Reporting Year" – SPD Data

Date/Time of Incident	Reporting Year
29/01/2012 11:04	2012
31/08/2012 21:00	2011
11/05/2017 07:21	2015

Table 5 Inconsistent Entries in "Date/Time of Incident" and "Reporting Year" – SPM Data

Date/Time of Incident	Reporting Year
14/04/2011 10:52	2010
31/01/2012 09:43	2012
17/02/2012 12:22	2012
03/03/2014 11:13	2014
27/03/2014 09:22	2014
24/04/2014 15:45	2013
16/05/2014 12:02	2013
07/03/2016 14:46	2013
01/04/2019 06:46	2018
28/04/2019 10:55	2018
03/05/2019 10:41	2018
14/03/2021 20:03	2017
23/03/2021 10:01	2021

Furthermore, as noted in Table 2, there are 66 categories recorded under "Main Cause" for SPD, which is one category more than the 65 categories recorded for SPM. This additional "Main Cause" is noted as follows; and over the 10-year period, only one fault in the district of Lanarkshire on 3rd March 2020 was identified by this "Main Cause".

- Failure of ehv/hv infeed from Adjacent District of Same PES Company

This is deemed insignificant and would not affect the investigation during the Discovery Phase, due to both the number of faults associated with this category (i.e., only one) and the project's focus on the categories closely related to weather conditions and events.

2.3.3 Other General Observations

It is observed in the historical network data for both SPD and SPM that faults categorised by "Deterioration due to Ageing or Wear (excluding corrosion)" account for the majority of the recorded faults, whilst faults categorised by "Cause Unknown" make up the second or third largest proportion of the recorded faults, as presented in Table 6. Additionally, there is also a category of "Causes Unclassified in this Table", to which 1,477 faults recorded in SPD and 152 faults recorded in SPM belong.

Table 6 Faults Categorised by "Deterioration due to Ageing or Wear (excl. Corrosion)" and "Cause Unknown"

Network	Total Number of Faults	Number of Faults	
		Deterioration due to Ageing or Wear (excl. Corrosion)	Cause Unknown
SPD	36,551	10,742 (or 29.39%)	7,200 (or 19.70%)
SPM	32,067	10,117 (or 31.55%)	4,387 (or 13.68%)

2.4 Summary of Available Data Quality

Following an overview of the historical network fault data provided by SPEN in Section 2.3, the data quality can be summarised in accordance with Table 1 as follows.

Table 7 Summary of Available Data Quality

Data Component	Available Data	Quality	Mitigation Measures
Timestamp for each network fault	Date/Time of Incident	The available data is a detailed timestamp with an accuracy down to a minute.	N/A
Record of the main cause of each network fault	Main Cause	The available data is a detailed description of the main cause of the network fault. However, it is important to understand the detailed principle for assigning a main cause, more specifically for those that may need a clear distinction.	This gap can be filled by some additional information that can be provided to improve clarity and by engaging with the responsible engineer(s) to gain an understanding of how the practice currently is.
Record of the geographical location of each network fault	– District – Location – Grid References X, Y	"District" identifies in which SPEN's district the network fault occurred. The link is provided with assistance from the geographical definition of each district. "Location" is short and vague, which cannot be linked to a geographical location without further processing. "Grid References X, Y" are not provided consistently, and errors exist in the record.	The option for filling the gap in Grid References X, Y may be using "Circuit Code", "HV Circuit Name" in combination with other available asset record. Another option is to improve the information provided in "Location", e.g., identifying the relevant post code area.
Record of the asset involved in each network fault	– Circuit Code – HV Circuit Name – HV Component	The information provided in "HV Component" describes the typical asset type that was involved in the network fault, whilst "Circuit Code" and "HV Circuit Name" can provide complement information with assistance from other relevant asset records. However, "unknown" is among the data entries under "HV Component", which requires clarification with regards to the principle for using this.	This gap can be filled by some additional information that can be provided to improve clarity and by engaging with the responsible engineer(s) to gain an understanding of how the practice currently is.
Record of the impact of each network fault	– Longest Stage Duration – CI – CHL	The available data include detailed CI and CHL assessed for the network fault, from the customer services' point of view. However, no information with regard to the scale of impact on the asset is available. This may help to understand the cost of restoration, and how preventative measures may be taken in light of the fault forecast.	Engagement with the relevant engineer(s) may be required to understand whether it is possible to provide this information or the necessity of the cost of restoration.

2.5 Analysis of SPD Historical Network Fault Data

This section aims to provide an analysis of the historical network fault data recorded for SPD from various points of view. This allows a detailed understanding of the fault data, which should help the investigation in the Discovery Phase.

2.5.1 Faults Categorised by "Main Cause"

There are 66 categories in total recorded under "Main Cause", and Table 8 presents the top 20 categories ranked by the number of faults. The top three weather-related categories include:

- Wind and Gale (excluding Windborne Material): 6,430 faults (17.59%)
- Lightning: 1,657 faults (4.53%)
- Snow, Sleet and Blizzard: 751 faults (2.05%)

Table 8 Top 20 Categories of "Main Cause" Ranked by the Number of Faults – SPD

Index	Main Cause	Number of Faults	%
1	Deterioration due to Ageing or Wear (excluding corrosion)	10,742	29.39%
2	Cause Unknown	7,200	19.70%
3	Wind and Gale (excluding Windborne Material)	6,430	17.59%
4	Lightning	1,657	4.53%
5	Causes Unclassified in this Table	1,477	4.04%
6	Birds (including Swans and Geese)	1,402	3.84%
7	Operational or Safety Restriction	989	2.71%
8	Snow, Sleet and Blizzard	751	2.05%
9	Growing or Falling Trees (not felled)	623	1.70%
10	Farm and Domestic Animals	561	1.53%
11	Extension of Fault Zone due to Fault Switching (including ASC held faults)	516	1.41%
12	Faulty Installation or Construction	431	1.18%
13	Wilful Damage or Interference	351	0.96%
14	Damage by Unknown Third Parties	274	0.75%
15	Damage by Other Third Parties	243	0.66%
16	Damage Involving Farm Workers or Farm Implements	243	0.66%
17	Rain	224	0.61%
18	Damage by Private Developers or their Contractors	219	0.60%
19	Vermin, Wild Animals and Insects	160	0.44%
20	Windborne Materials	149	0.41%
	Others	1,909	5.22%
	Total	36,551	100%

2.5.2 Seasonal Trend of Top Weather-related Faults

The seasonal trend of faults recorded against the weather-related categories in Table 8 is analysed in this subsection. More specifically, the focus is on the following five weather-related categories.

- Wind and Gale (excluding Windborne Material)
- Lightning
- Snow, Sleet and Blizzard
- Rain
- Windborne Material

2.5.2.1 Wind and Gale (excluding Windborne Material)

Wind and gale (excluding windborne material) is the most significant weather-related cause for network faults, accounting for 17.59% of the faults recorded for the decade between 2010 and 2020. As illustrated in Figure 1, the monthly number of faults exhibits a seasonal trend, i.e., higher number of faults normally occurs in the months between October and February. Nonetheless, it is observed that the number of faults presents a declining trend from 2010 to 2020. This may need to be further investigated in the Discovery Phase with regards to how this trend correlates to the historical weather data.

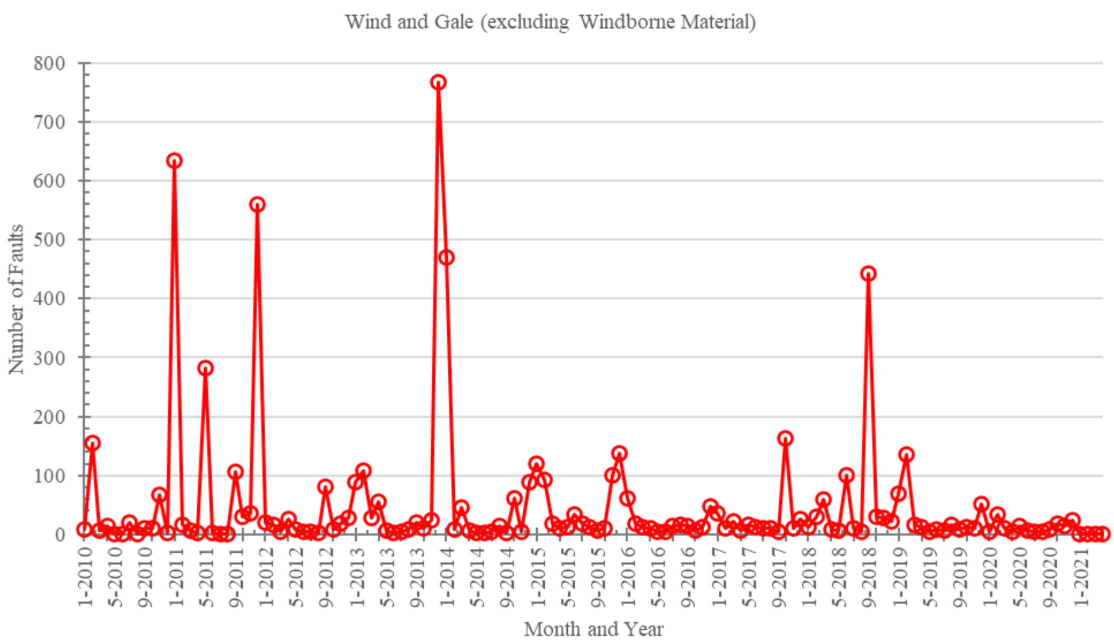


Figure 1: Monthly Number of Faults Caused by Wind and Gale (excluding Windborne Material) from 2010 to 2020

On the other hand, it is advised by SPEN that 76 faults and over within a window of 24 hours can qualify for an exceptional event. Therefore, Figure 2 presents the statistics for the number of faults per month over the 10-year period from 2010 to 2020.

It can be seen in Figure 2 that 117 months had 75 or less faults whilst 19 months had 76 or more faults. This implies that the majority of faults caused by Wind and Gale (excluding Windborne Material) were not able to be deemed as an exceptional event.

Additionally, the statistics also show that 75% of the months from 2010 to 2020 had 30 or less faults caused by Wind and Gale (excluding Windborne Material), whereas 50% of the months had between 6 and 30 such faults. This may imply that there may be a tangible benefit to SPEN should the number of such faults can be forecasted, not only for severe storms but also for more frequent adverse weather conditions.

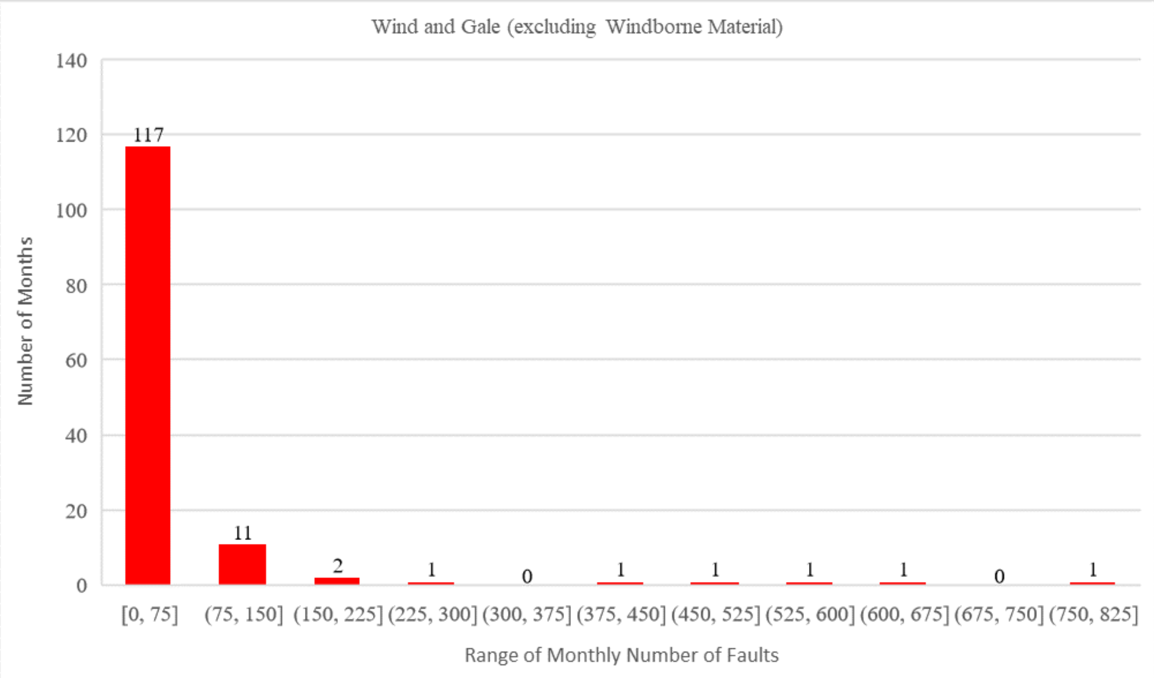


Figure 2: Statistics for Monthly Number of Faults Caused by Wind and Gale (excluding Windborne Material)

2.5.2.2 Lightning

Lightning is the second most significant weather-related cause for network faults, accounting for 4.59% of the faults recorded for the decade between 2010 and 2020. As illustrated in Figure 3, the monthly number of faults also demonstrates a seasonal trend, i.e., higher number of faults normally occurs in the months between May and September. This seasonal trend can be consistently observed over the 10-year period from 2010 to 2020. However, it is raised by the Met Office team that lightning forecast is difficult particularly for a 5-day window.

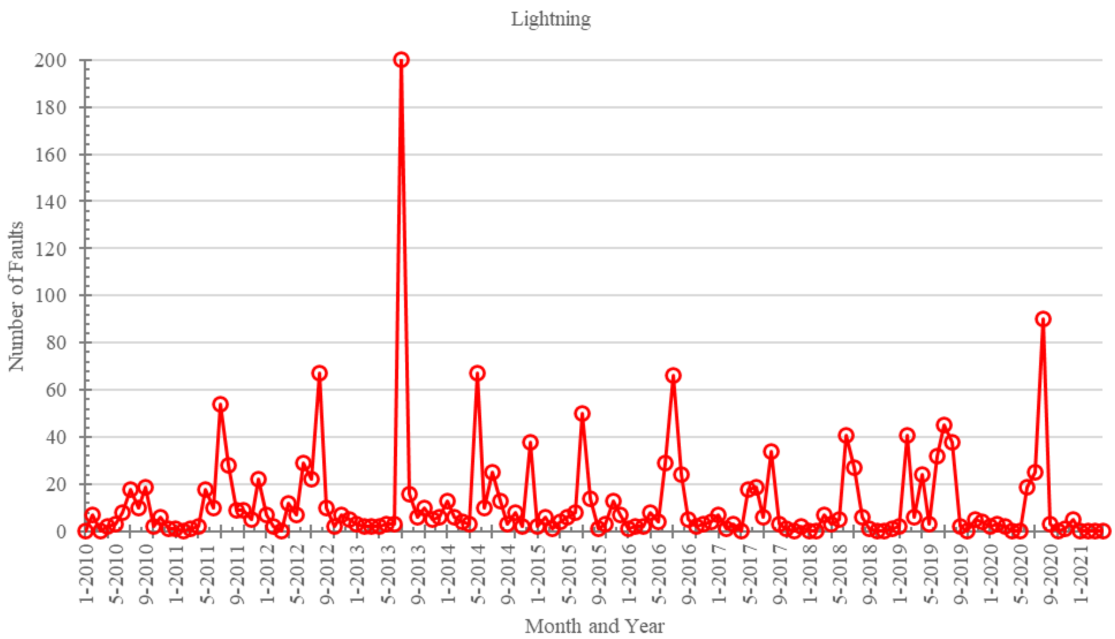


Figure 3: Monthly Number of Faults Caused by Lightning from 2010 to 2020

Figure 4 shows the statistics for the number of faults per month over the 10-year period from 2010 to 2020. It is found that 134 months had no more than 75 faults whilst one month had 90 faults and another one had 200 faults. This implies that in general faults caused by lightning would not constitute an exceptional event.

Further statistics show that 75% of the months from 2010 to 2020 had 13 or less faults caused by Lightning, whereas 50% of the months had between 2 and 13 such faults.

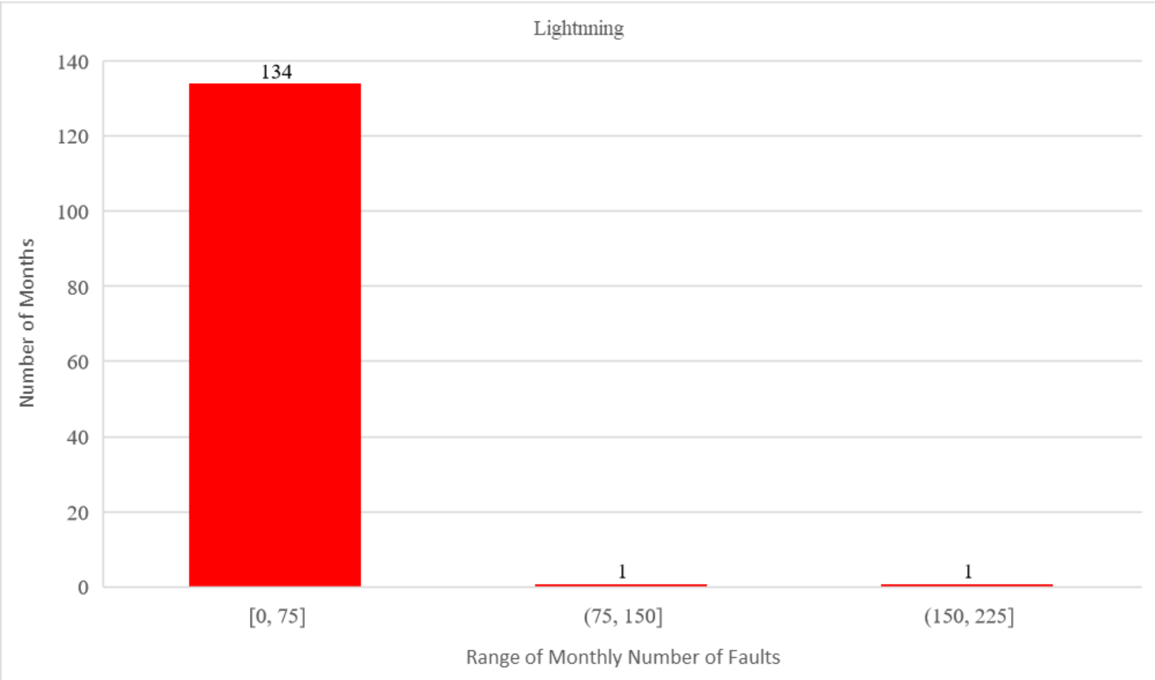


Figure 4: Statistics for Monthly Number of Faults Caused by Lightning

2.5.2.3 Snow, Sleet and Blizzard

The third weather-related cause for network faults is "Snow, Sleet and Blizzard", and the recorded number of faults per month is illustrated in Figure 5. As expected, the faults caused by Snow, Sleet and Blizzard would occur in November the earliest and likely continue to be observed until April in the next year.

The monthly number of faults would not exceed 50 apart from November 2010, and March and April 2012. Similar to wind and gale related faults, majority of faults caused by Snow, Sleet and Blizzard would not constitute an exceptional event; and the ability to forecast the number of faults that could be brought by Snow, Sleet and Blizzard should be able to improve network performance.

Unlike Figure 2 and Figure 4, the distribution shown in Figure 6 adopts a bin size of 10 faults per month for the histogram due to the overall low number of faults per month. According to Figure 6, 121 months had less than 10 faults, and in fact further statistics show that 75% of the months would have one or no fault due to Snow, Sleet and Blizzard. This is expected in light of the seasonal nature of these faults. This therefore also means that the forecast for this category of faults should be more focused on specific storms in order to maximise the benefit provided to the Control Room. Additionally, the location and magnitude of impact would be particularly important.

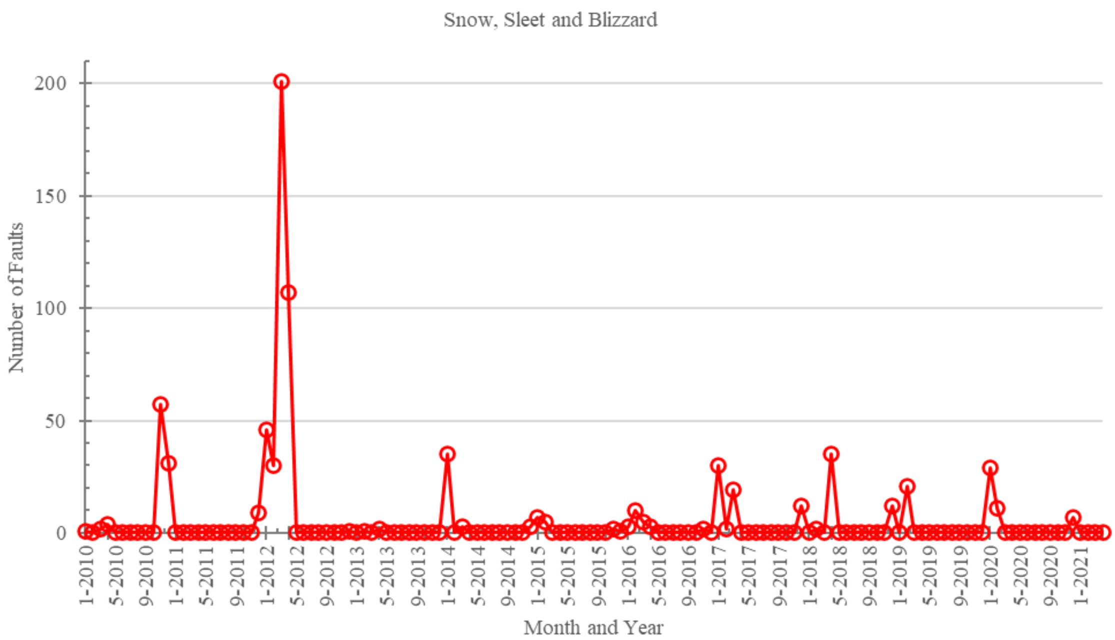


Figure 5: Monthly Number of Faults Caused by Snow, Sleet and Blizzard from 2010 to 2020

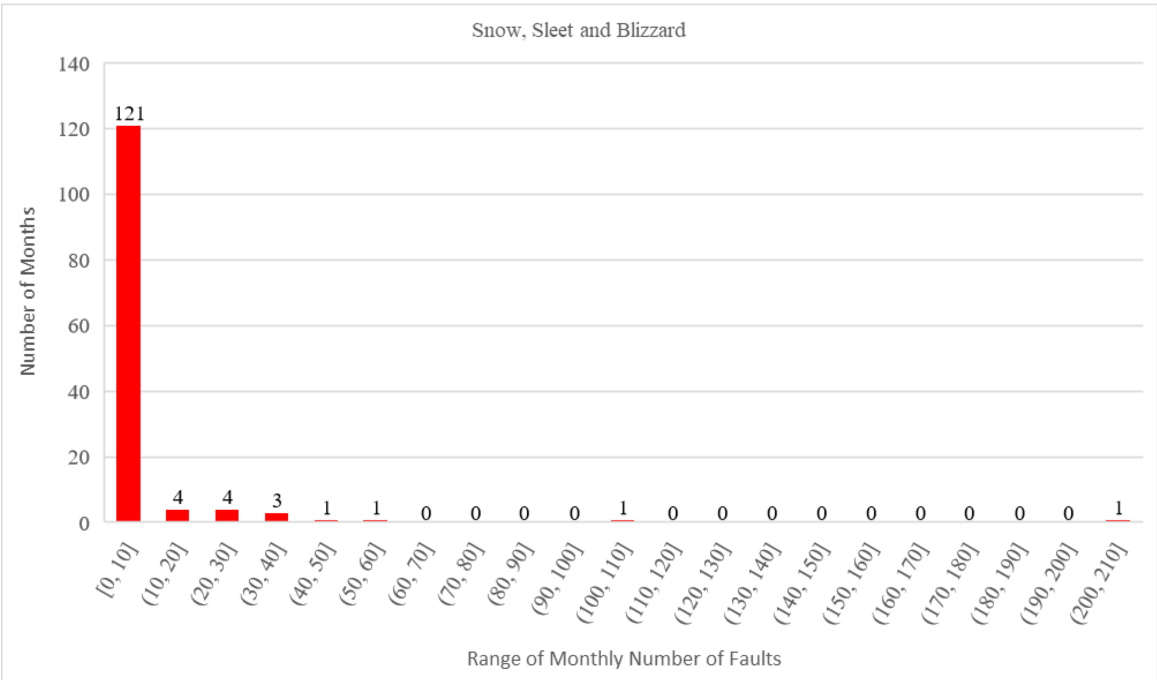


Figure 6: Statistics for Monthly Number of Faults Caused by Snow, Sleet and Blizzard

2.5.2.4 Rain

The total number of faults caused by Rain is 224 over the 10-year period under analysis, accounting for 0.61% of the total recorded faults. The number of faults per month is presented in Figure 7. It is observed that there are only fewer faults recorded in 2010 and 2011, compared with the other years. On the other hand, late 2013 and early 2014, late 2015, and mid 2016 saw higher number of faults than the other months over the 10-year period. This may need to be further investigate with the corresponding historical weather data.

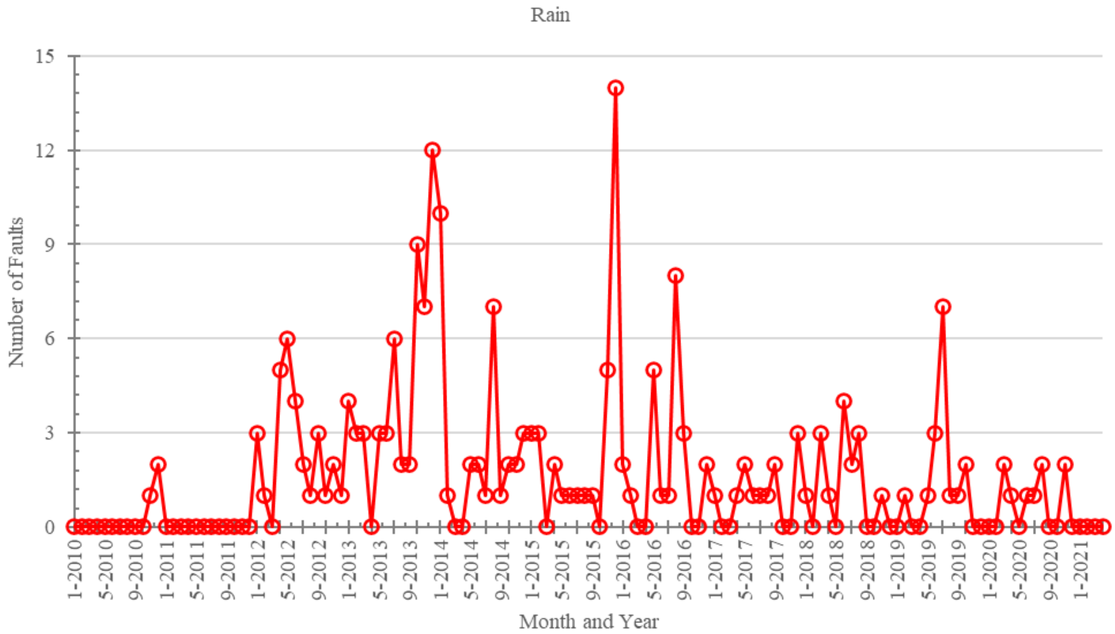


Figure 7: Monthly Number of Faults Caused by Rain from 2010 to 2020

The distribution shown in Figure 8 adopts a much smaller bin size of 2 faults per month for the histogram, compared with the histograms presented previously. This is primarily driven by the low number of faults per month. According to Figure 8, 106 months had 2 or less faults caused by Rain, which is more than 75% of the months. This means that the faults caused by Rain should not only be assessed by the number of faults but also the magnitude of their impact (e.g., measured by number of affected customers and the duration of these interruptions). This therefore also means that the forecast for this category of faults should be more focused on specific rainfall events in order to maximise the benefit provided to the Control Room. Further, the location and magnitude of impact would be particularly important.

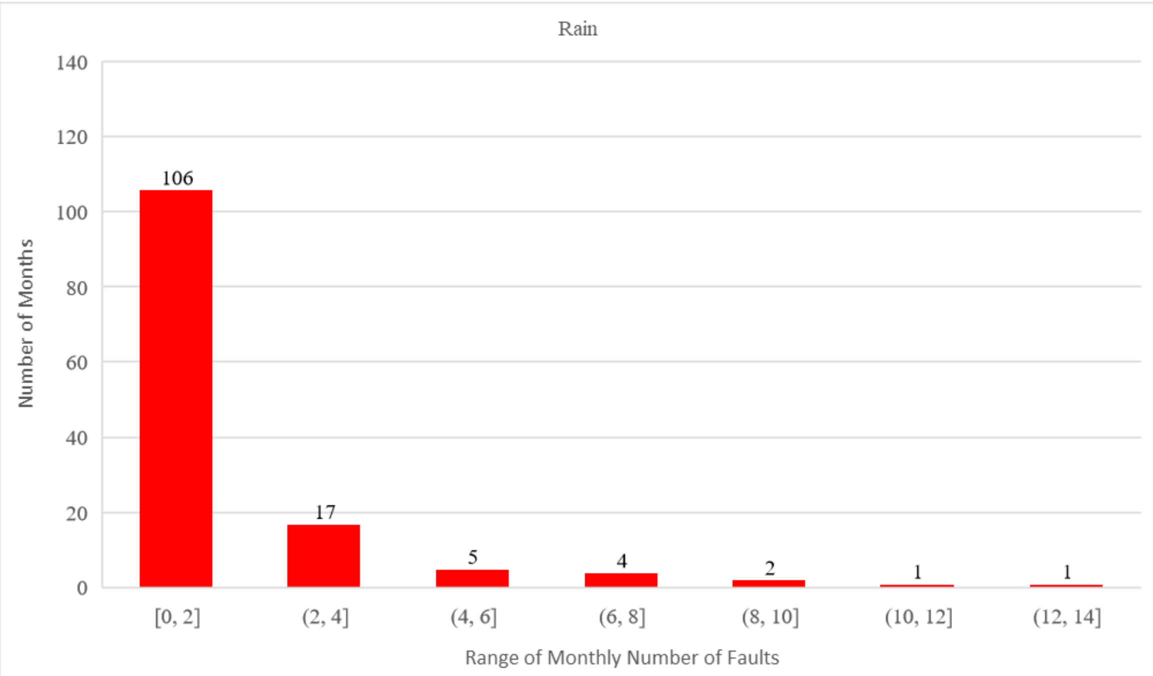


Figure 8: Statistics for Monthly Number of Faults Caused by Rain

2.5.2.5 *Windborne Material*

Windborne Material is a separate category of "Main Cause" for the recorded faults, which has a total number of 149 faults recorded against this category over the 10-year period. Figure 9 shows that the number of faults per month varies but mostly remains between 1 and 4 faults per month throughout the 10-year period. This is a different trend as exhibited in Figure 1, i.e., less seasonal variation is observed in Figure 9 comparing to Figure 1. The distribution shown in Figure 10 also adopts a bin size of 2 faults per month for the histogram. This is also driven by the low number of faults per month. According to Figure 10, 120 months had 2 or less faults caused by Windborne Material, and the highest number of faults per month is 8. Further, it is worth investigating how faults caused by Windborne Material correlate to those presented in Section 2.5.2.1.

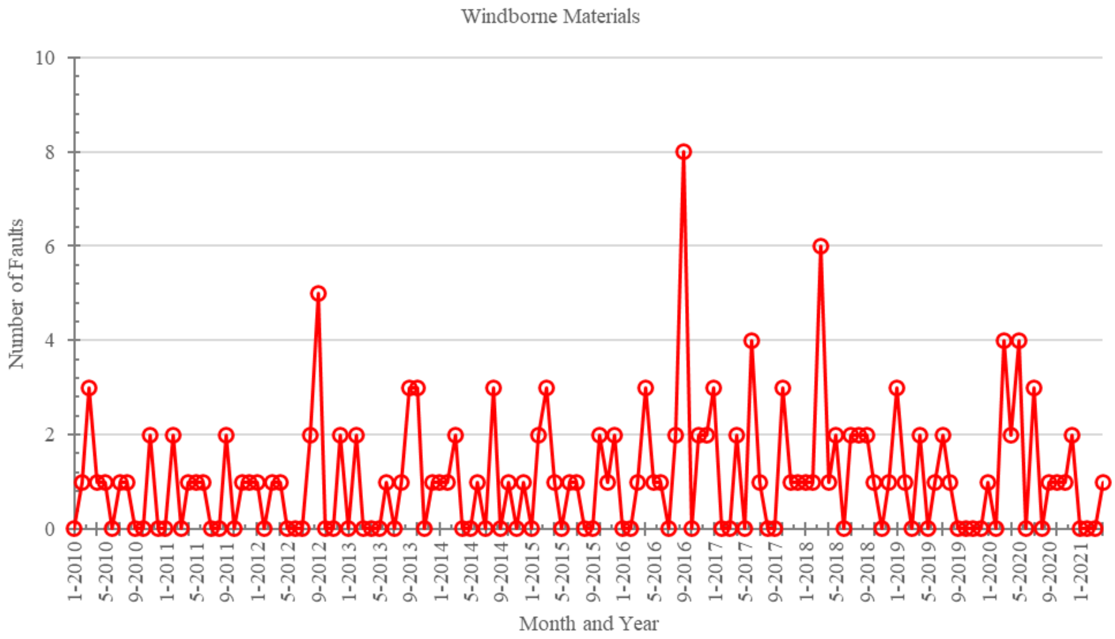


Figure 9: Monthly Number of Faults Caused by Windborne Material from 2010 to 2020

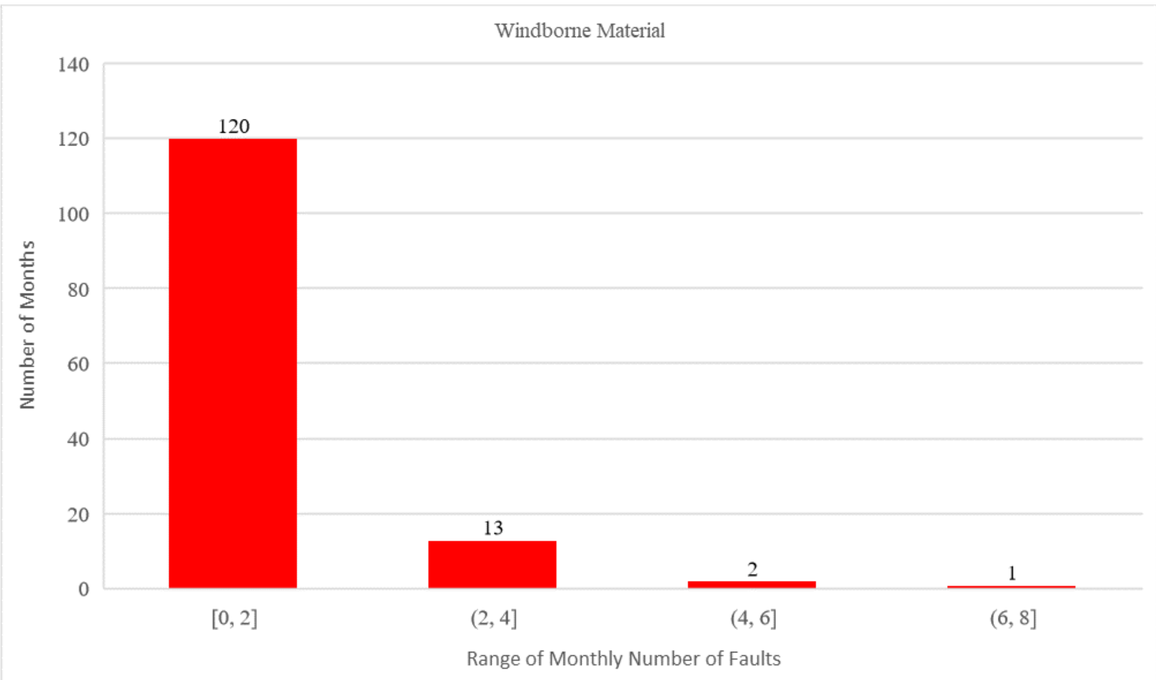


Figure 10: Statistics for Monthly Number of Faults Caused by Windborne Material

2.5.3 Geographical Distribution of Top Weather-related Faults

Prior to analysing the five top weather-related faults, the numbers of all the faults recorded in each of the seven SPD districts are shown in Figure 11. The district of Central and Fife, Ayrshire and Clyde South, and Dumfries each has over 7000 faults recorded; in other words, each of these three districts accounts for close to 20% of the total faults recorded for SPD. The district of Glasgow and Clyde North has the lowest number of faults recorded.

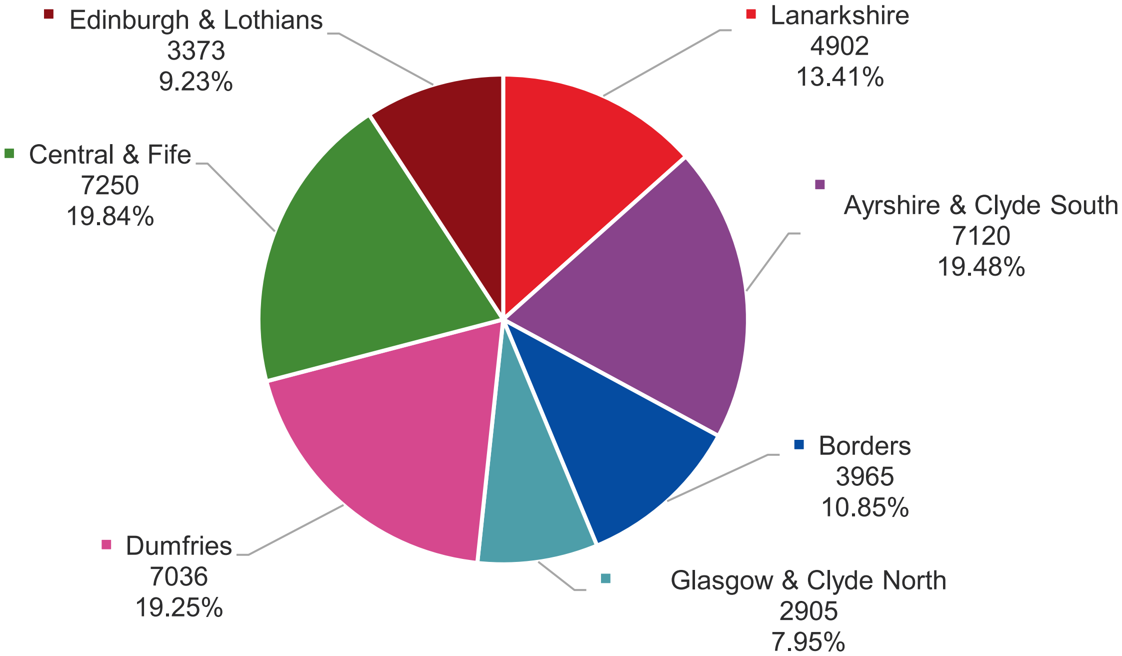


Figure 11: Number of Faults Recorded in Each of the SPD Districts

2.5.3.1 Wind and Gale (excluding Windborne Material)

According to Figure 12, the district of Dumfries has the highest number of faults caused by Wind and Gale (excluding Windborne Material), followed by the district of Ayrshire and Clyde South, and Central and Fife. This is consistent with the observation in Figure 11, i.e., these three districts account for over 60% of the faults caused by Wind and Gale (excluding Windborne Material).

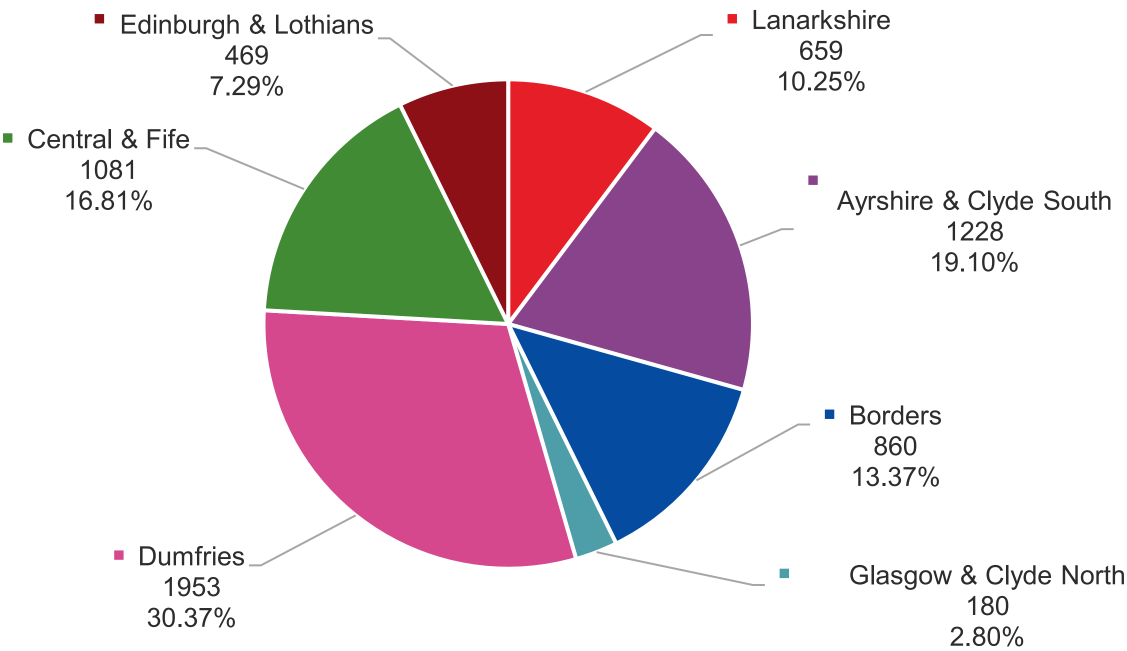


Figure 12: Number of Faults Caused by Wind and Gale (excluding Windborne Material) in Each of the SPD Districts

2.5.3.2 *Lightning*

Figure 13 shows a similar distribution of faults caused by Lightning across the seven districts comparing to that in Figure 11 and Figure 12. The district of Dumfries has the highest number of faults, i.e., 428 faults (25.83% of total faults caused by Lightning). The district of Central and Fife, and the district of Ayrshire and Clyde South has 327 faults (19.73%) and 306 faults (18.47%), respectively.

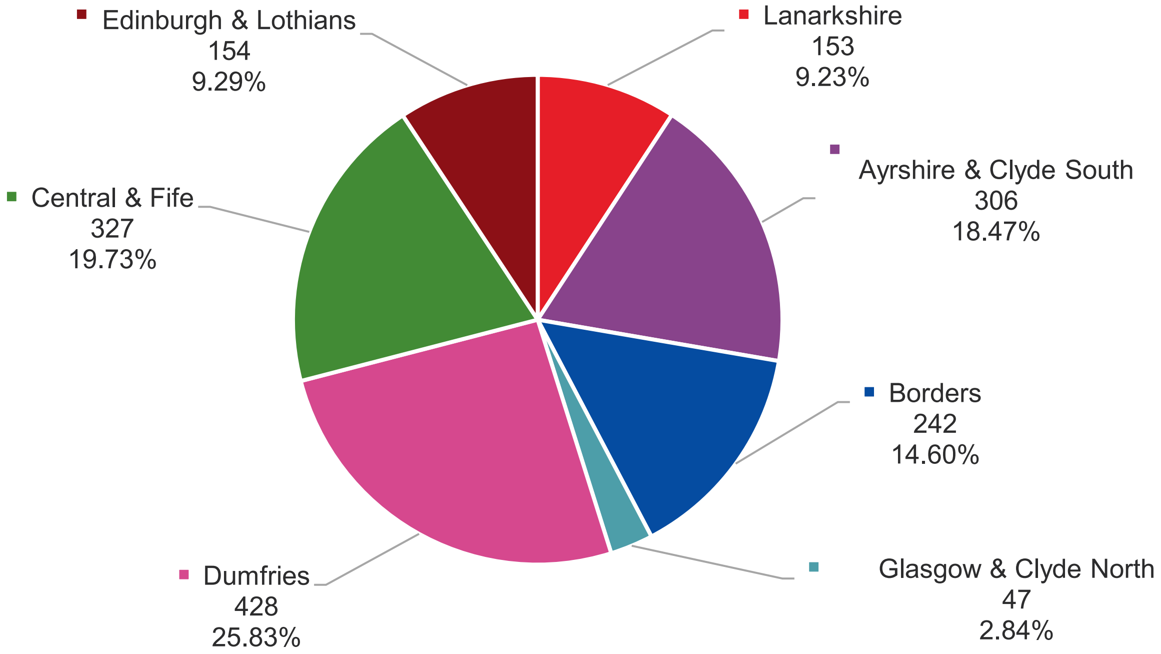


Figure 13: Number of Faults Caused by Lightning in Each of the SPD Districts

2.5.3.3 *Snow, Sleet and Blizzard*

The districts of Dumfries, and Central and Fife have 243 faults and 150 faults, respectively. They account for over half of the total faults caused by Snow, Sleet and Blizzard. Unlike faults caused by Wind and Gale (excluding Windborne Material) or Lightning, the number of faults caused by Snow, Sleet and Blizzard recorded in the district of Ayrshire and Clyde South is limited, i.e., 63 faults or only 8.39% of such faults. The district of Borders has 161 faults due to Snow, Sleet and Blizzard, which is the third highest.

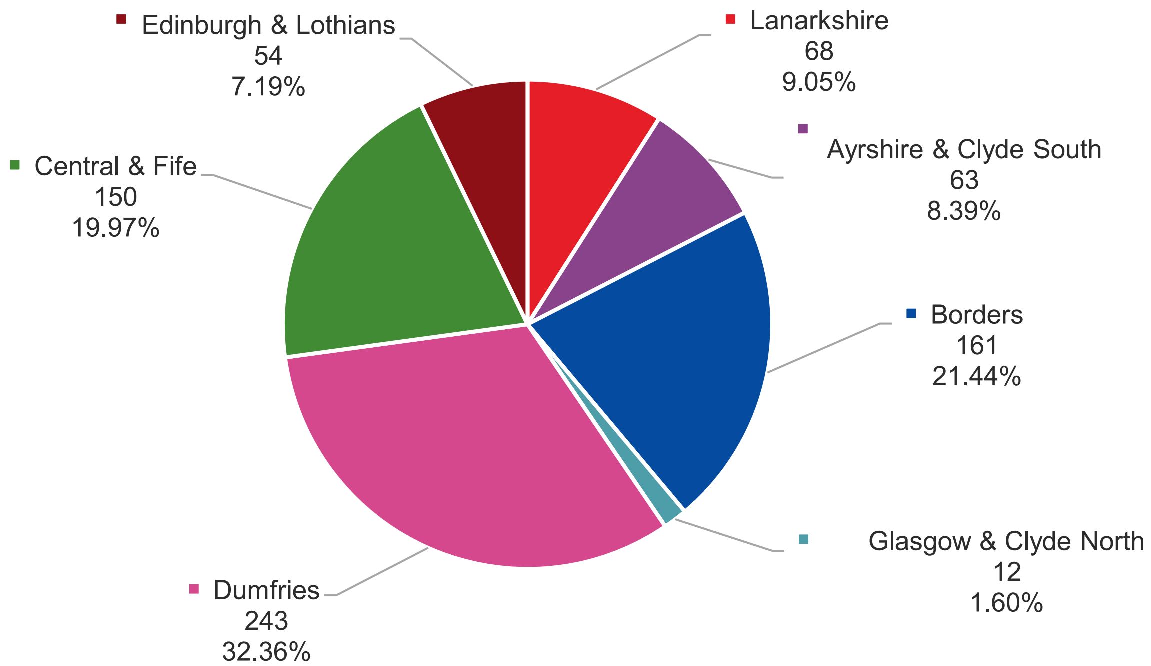


Figure 14: Number of Faults Caused by Snow, Sleet and Blizzard in Each of the SPD Districts

2.5.3.4 Rain

The highest number of faults caused by Rain is in the district of Dumfries, which is followed by the districts of Borders, and Central and Fife.

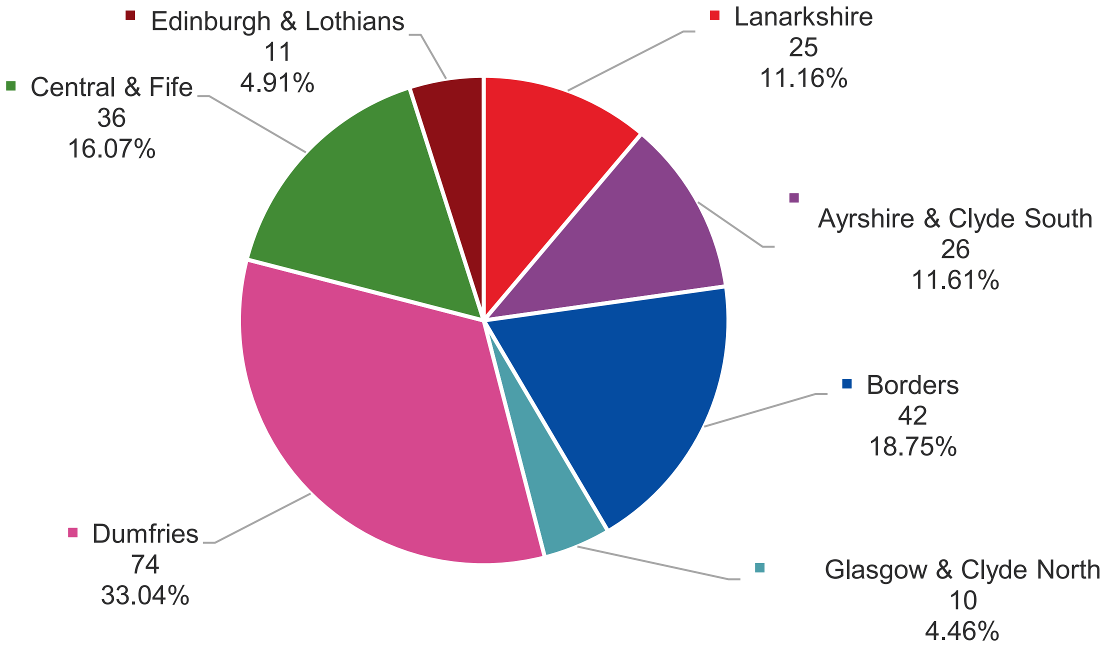


Figure 15: Number of Faults Caused by Rain in Each of the SPD Districts

2.5.3.5 Windborne Material

Each of the districts including Dumfries, Central and Fife, Borders, and Ayrshire and Clyde South have 30 faults caused by Windborne Material, or thereabouts. The distribution of this category of faults across the seven districts is mostly consistent with that observed in Figure 12.

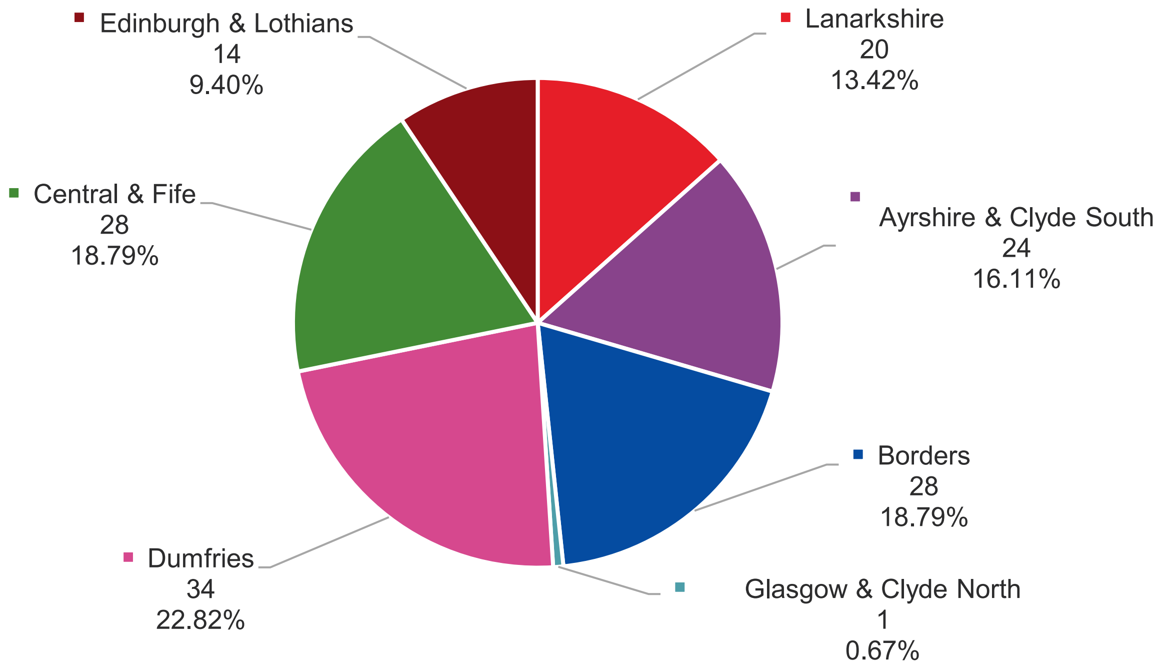


Figure 16: Number of Faults Caused by Windborne Material in Each of the SPD Districts

2.5.4 Impact of Top Weather-related Faults

This subsection presents the impact of the five top weather-related faults based on the longest duration, Customer Interruptions (CI), and Customer Hours Lost (CHL) provided in the historical network fault data. In order to analyse their impact, the box and whisker chart (also known as boxplot) is used, which is able to illustrate the distribution of the recorded data.

More specifically, the bottom and top boundaries of the box represent the first and third quartiles, respectively; and the line in the middle of the box is the median value. Meanwhile, the lower whisker corresponds to the minimum value, and the higher whisker is the boundary for outliers, which are represented by dots in the figures. The cross mark is the average value of the relevant data.

2.5.4.1 Wind and Gale (excluding Windborne Material)

The distribution of Duration Longest, CI and CHL is presented in Figure 17. Due to the long tail distribution of these metrics (i.e., the number of outliers is significant), the outliers shown in Figure 17 do not included all the outliers. This long tail distribution also leads to the average values of these metrics being skewed, compared with the corresponding median values. The maximum values of these metrics are as follows.

- Maximum of Duration Longest is 148.8 hours
- Maximum CI is 16,011
- Maximum CHL is 70,004

According to Figure 17, 75% of the Duration Longest is less than 10 hours and 50% of the Duration Longest is less than 4 hours, whereas the Duration Longest between 1 hour and 10 hours accounts for half of the faults recorded. Three quarters of the faults resulted in a CI less than 140 and half of them had a CI less than 30; whilst three quarters of the faults resulted in a CHL less than 400 and half of them had a CHL less than 100. On the other hand, half of the faults show a range of CI between 7 and 140, and a range of CHL between 20 and 400.

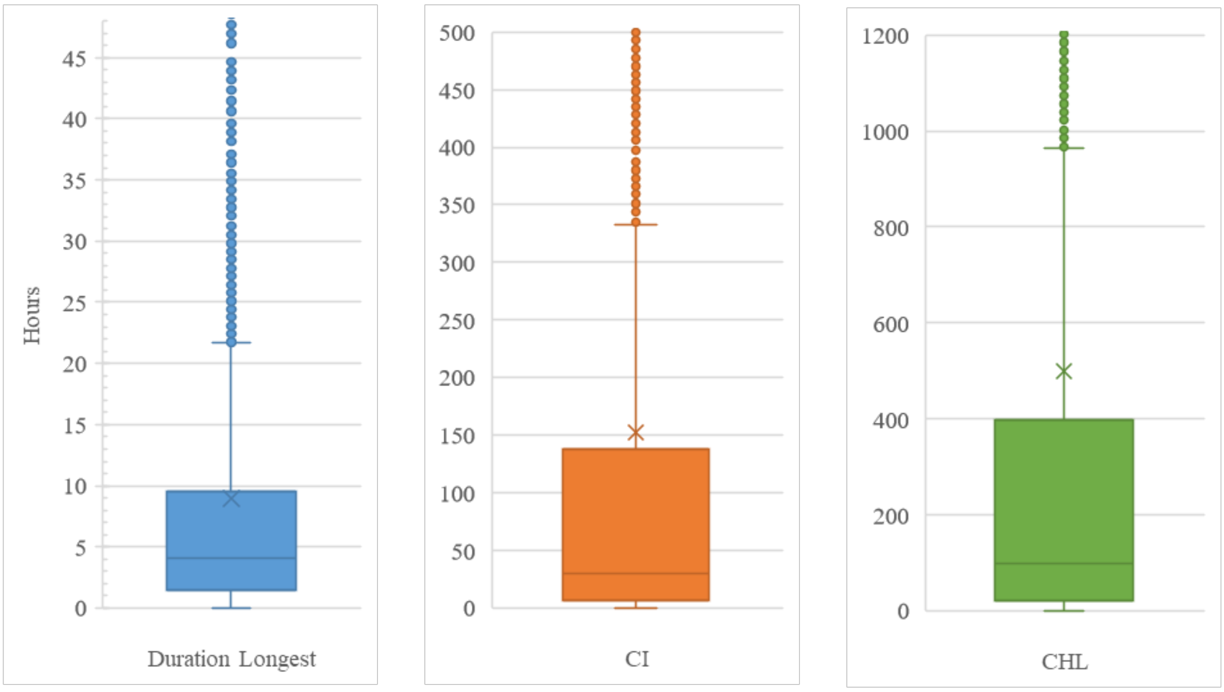


Figure 17: Boxplot for Duration Longest, CI and CHL of the Faults Caused by Wind and Gale (excl. Windborne Material)

2.5.4.2 *Lightning*

The magnitude of Lightning faults' impact is demonstrated in Figure 18. Similarly, the outliers are not all shown in Figure 18 due to the long tail distributions. The maximum values recorded for these metrics are as follows.

- Maximum of Duration Longest is 55.2 hours
- Maximum CI is 16,837
- Maximum CHL is 33,060.25

According to Figure 18, 75% of the Duration Longest is less than 5 hours and 50% of the Duration Longest is less than 2.5 hours, whereas, the Duration Longest between 0.38 and 5 hours accounts for half of the faults recorded. Three quarters of the faults resulted in a CI less than 150 and half of them had a CI less than 27; whilst three quarters of the faults resulted in a CHL less than 200 and half of them had a CHL less than 20. On the other hand, half of the faults show a range of CI between 5 and 150, and a range of CHL between 10 and 200.

Unlike the faults caused by Wind and Gale (excluding Windborne Material), the faults caused by Lightning would have less impact from the Duration Longest and CHL's points of view, i.e., the faults were cleared and restored faster. Nonetheless, the number of affected customers, as represented by CI, exhibits a similar distribution.

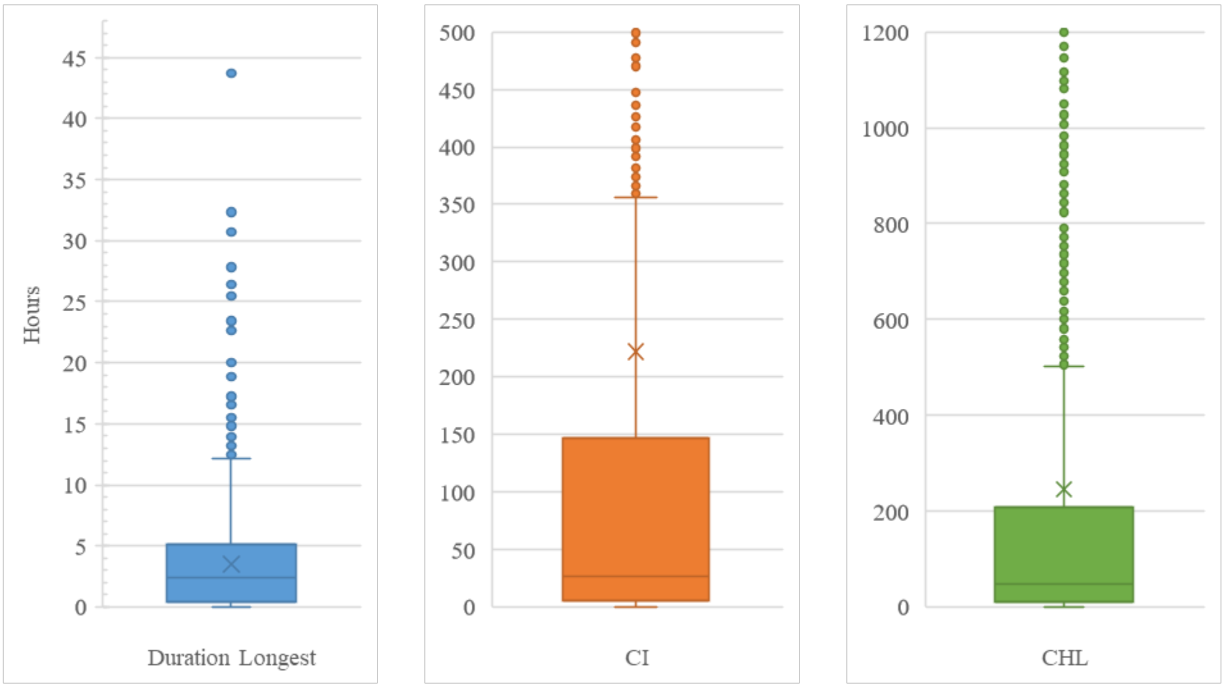


Figure 18: Boxplot for Duration Longest, CI and CHL of the Faults Caused by Lightning

2.5.4.3 *Snow, Sleet and Blizzard*

The distribution of Duration Longest, CI and CHL is presented in Figure 19. The long tail distribution of these metrics (i.e., the number of outliers is significant) lead to the outliers shown in Figure 19 not including all the outliers. As mentioned previously, this long tail distribution skews the average values of these metrics, compared with the corresponding median values. The maximum values of these metrics are as follows.

- Maximum of Duration Longest is 189.2 hours
- Maximum CI is 6,609
- Maximum CHL is 85,665.7

According to Figure 19, 75% of the Duration Longest is less than 9 hours and 50% of the Duration Longest is less than 1.5 hours, whereas, the Duration Longest between 1.5 hour and 9 hours accounts for half of the faults recorded. Three quarters of the faults resulted in a CI less than 170 and half of them had a CI less than 40; whilst three quarters of the faults resulted in a CHL less than 450 and half of them had a CHL less than 130. On the other hand, half of the faults show a range of CI between 9 and 166, and a range of CHL between 29 and 445.

It is observed that the Duration Longest and CHL is much higher than those of the faults caused by either Wind and Gale (excluding Windborne Material) or Lightning, whilst the CI is much lower.

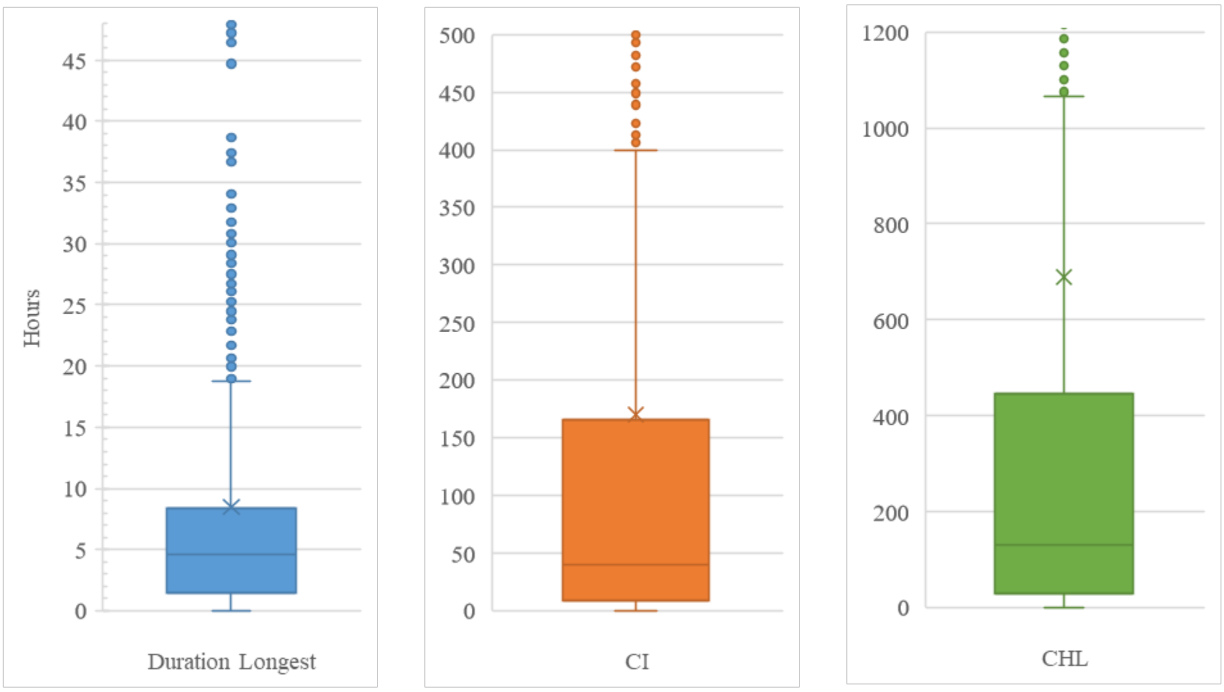


Figure 19: Boxplot for Duration Longest, CI and CHL of the Faults Caused by Snow, Sleet and Blizzard

2.5.4.4 Rain

Similar to previous figures, Figure 20 does not show all the outliers for CI and CHL due to their long tail distributions. However, the distribution of the Duration Longest includes all the recoded data. The maximum values of these metrics are as follows.

- Maximum of Duration Longest is 23.6 hours
- Maximum CI is 1,351
- Maximum CHL is 6,205.1

According to Figure 20, 75% of the Duration Longest is less than 4.5 hours and 50% of the Duration Longest is less than 2.5 hours, whereas, the Duration Longest between 1 hour and 4.5 hours accounts for half of the faults recorded. Three quarters of the faults resulted in a CI less than 185 and half of them had a CI less than 45; whilst three quarters of the faults resulted in a CHL less than 240 and half of them had a CHL less than 78. On the other hand, half of the faults show a range of CI between 13 and 184, and a range of CHL between 22 and 236.

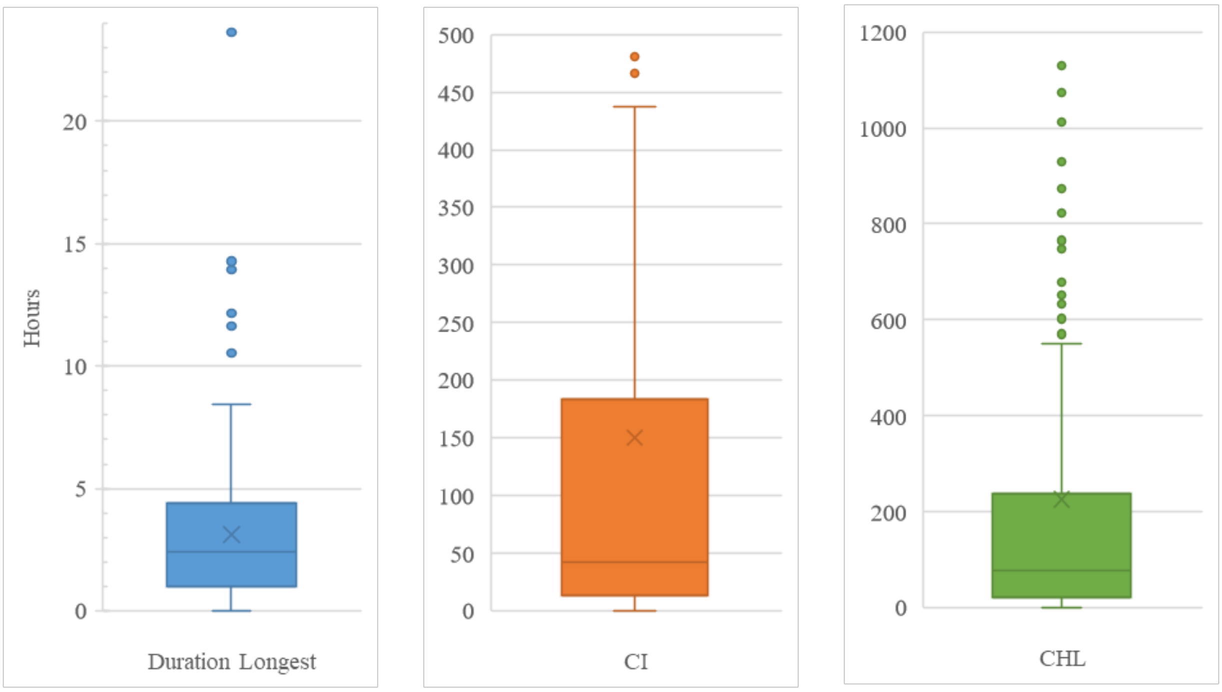


Figure 20: Boxplot for Duration Longest, CI and CHL of the Faults Caused by Rain

2.5.4.5 Windborne Material

Similarly, Figure 21 does not show all the outliers for CI and CHL due to their long tail distributions, whilst it presented the whole distribution of the Duration Longest including all the recoded data. The maximum values of these metrics are as follows, and it should be noted that a substantial gap between the maximum value and the next highest value is observed for both recoded CI and CHL values.

- Maximum of Duration Longest is 20.4 hours
- Maximum CI is 9,463 (followed by 1,277)
- Maximum CHL is 3,789.3 (followed by 1,012.3)

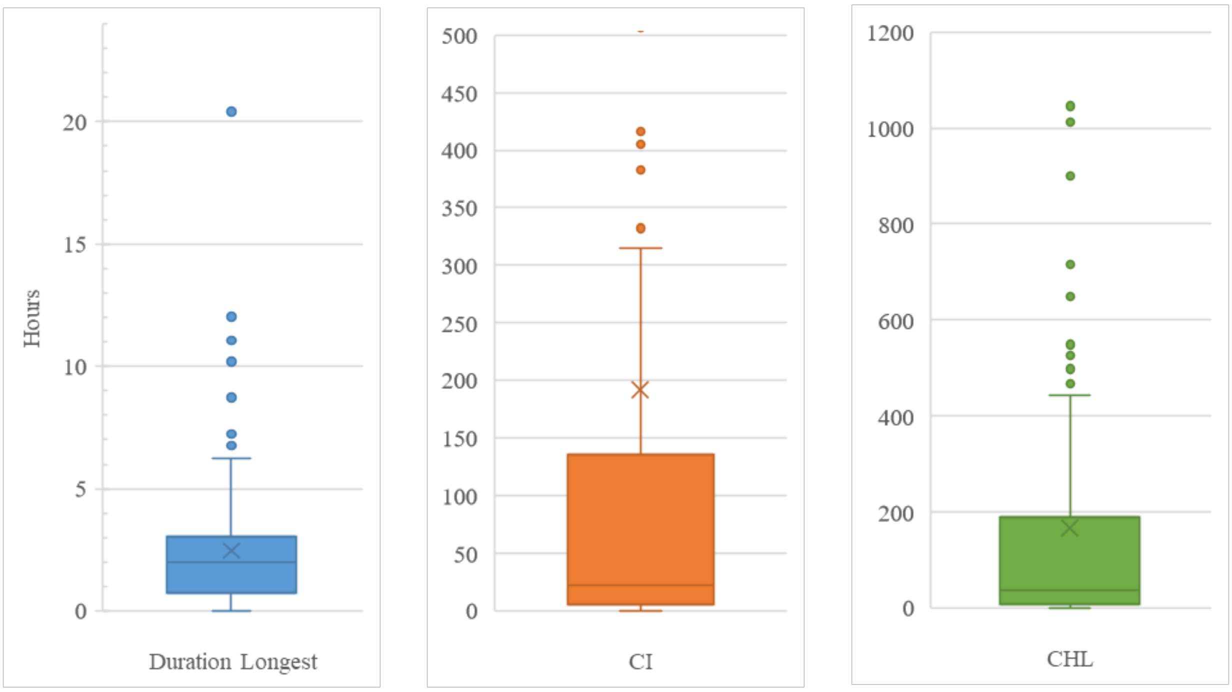


Figure 21: Boxplot for Duration Longest, CI and CHL of the Faults Caused by Windborne Material

According to Figure 21, 75% of the Duration Longest is less than 3 hours and 50% of the Duration Longest is less than 2 hours, whereas, the Duration Longest between 0.7 and 3 hours accounts for half of the faults recorded. Three quarters of the faults resulted in a CI less than 136 and half of them had a CI less than 22; whilst three quarters of the faults resulted in a CHL less than 188 and half of them had a CHL less than 38. On the other hand, half of the faults show a range of CI between 6 and 136, and a range of CHL between 7 and 188.

The faults caused by Windborne Material had a much less significant impact with regard to their durations, which are approximately half of the durations of the faults caused by Wind and Gale (excluding Windborne Material). However, from the CI’s point of view, half of the faults either caused by Windborne Material or by Wind and Gale (excluding Windborne Material) would lead to a CI in the range between 5 and 140.

2.6 Analysis of SPM Historical Network Fault Data

This section aims to provide an analysis of the historical network fault data recorded for SPM from various points of view. This allows a detailed understanding of the fault data, which should help the investigation in the Discovery Phase.

2.6.1 Faults Categorised by "Main Cause"

There are 65 categories in total recorded under "Main Cause", and Table 9 presents the top 20 categories ranked by the number of faults. The top three weather-related categories include:

- Wind and Gale (excluding Windborne Material): 4,638 faults (14.46%)
- Lightning: 1,205 faults (3.76%)
- Snow, Sleet and Blizzard: 305 faults (0.95%)

Table 9 Top 20 Categories of "Main Cause" Ranked by the Number of Faults – SPM

Index	Main Cause	Number of Faults	%
1	Deterioration due to Ageing or Wear (excluding corrosion)	10,117	31.55%
2	Wind and Gale (excluding Windborne Material)	4,638	14.46%
3	Cause Unknown	4,387	13.68%
4	Extension of Fault Zone due to Fault Switching (including ASC held faults)	1,938	6.04%
5	Growing or Falling Trees (not felled)	1,475	4.60%
6	Birds (including Swans and Geese)	1,427	4.45%
7	Lightning	1,205	3.76%
8	Faulty Installation or Construction	903	2.82%
9	Faulty Manufacturing, Design, Assembly or Materials	679	2.12%
10	Operational or Safety Restriction	476	1.48%
11	Vermin, Wild Animals and Insects	349	1.09%
12	Farm and Domestic Animals	318	0.99%
13	Snow, Sleet and Blizzard	305	0.95%
14	Damage Involving Farm Workers or Farm Implements	302	0.94%
15	Damage by Unknown Third Parties	298	0.93%
16	Windborne Materials	245	0.76%
17	Inadequate or Faulty Maintenance	222	0.69%
18	Wilful Damage or Interference	211	0.66%
19	Damage by Private Developers or their Contractors	181	0.56%
20	Unsuitable Protection Characteristics	171	0.53%
	Others	2,220	6.92%
	Total	32,067	100%

2.6.2 Seasonal Trend of Top Weather-related Faults

The seasonal trend of faults recorded against the weather-related categories in Table 9 is analysed in this subsection. More specifically, the focus is on the following five weather-related categories.

- Wind and Gale (excluding Windborne Material)
- Lightning
- Snow, Sleet and Blizzard
- Windborne Material

2.6.2.1 Wind and Gale (excluding Windborne Material)

Wind and gale (excluding windborne material) is the most significant weather-related cause for network faults, accounting for 14.46% of the faults recorded for the decade between 2010 and 2020. As illustrated in Figure 22, the monthly number of faults exhibits a seasonal trend, i.e., higher number of faults normally occurs in the months between October and February. Additionally, it is observed that the number of faults presents a declining trend after 2015. This may need to be further investigated in the Discovery Phase with regards to how this trend correlates to the historical weather data.

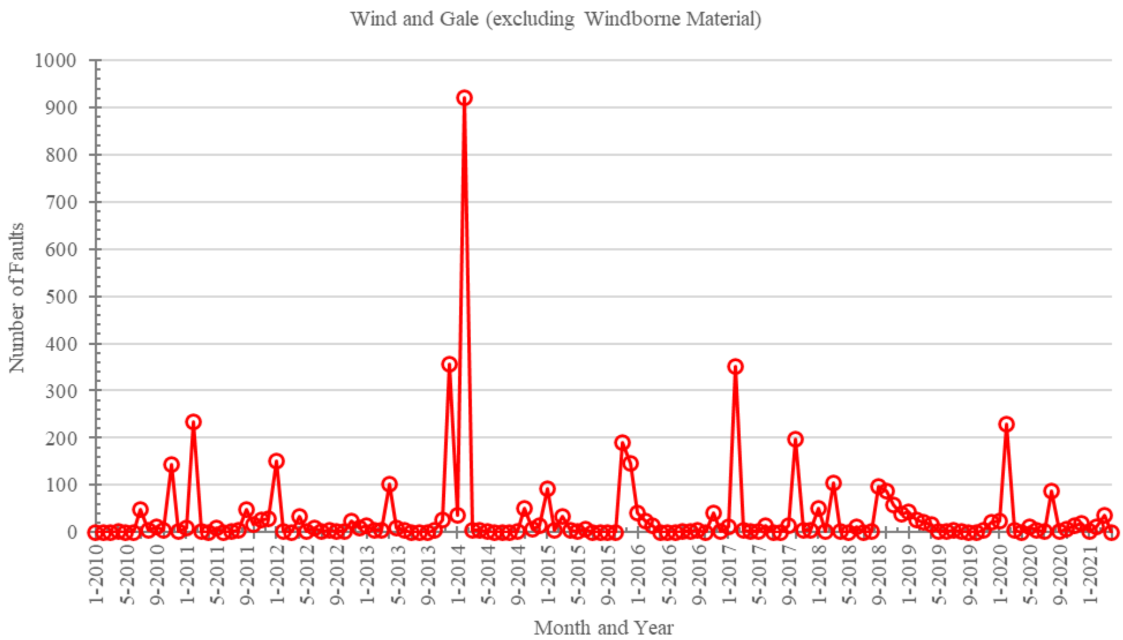


Figure 22: Monthly Number of Faults Caused by Wind and Gale (excluding Windborne Material) from 2010 to 2020

It can be seen in Figure 23 that 120 months had 75 or less faults whilst 16 months had 76 or more faults. This implies that the majority of faults caused by Wind and Gale (excluding Windborne Material) were not able to be deemed as an exceptional event.

Additionally, the statistics also show that 75% of the months from 2010 to 2020 had 24 or less faults caused by Wind and Gale (excluding Windborne Material), whereas 50% of the months had between 1 and 24 such faults. This may imply that there may be a tangible benefit to SPEN should the number of such faults can be forecasted, not only for severe storms but also for more frequent adverse weather conditions.

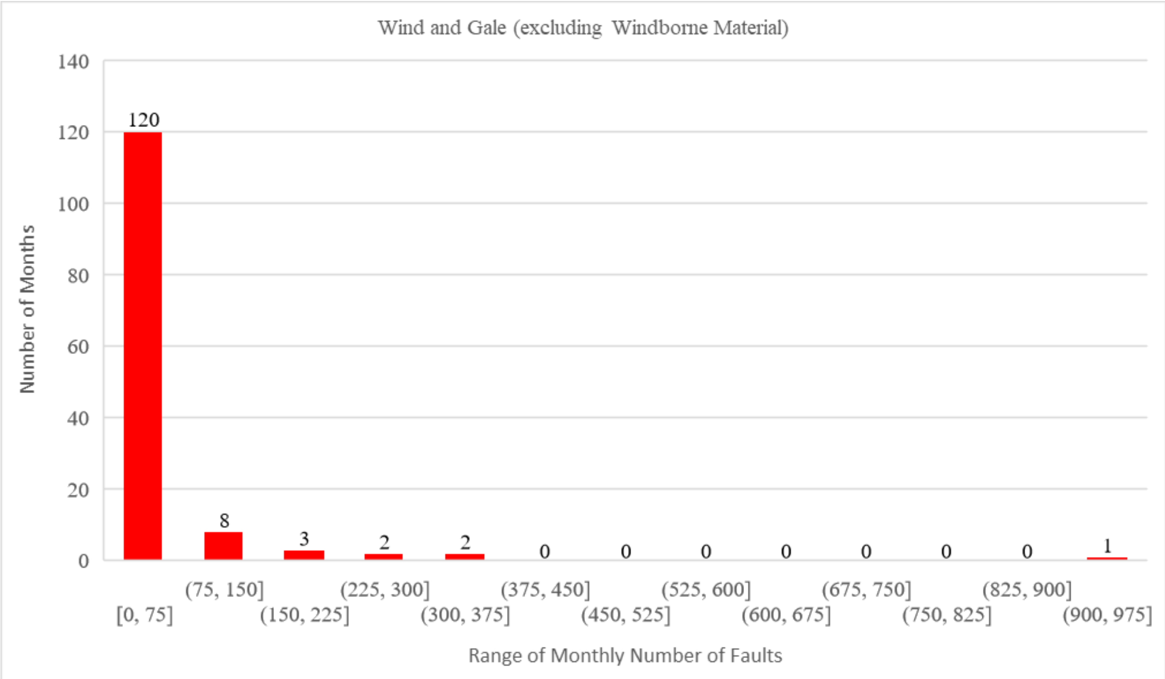


Figure 23: Statistics for Monthly Number of Faults Caused by Wind and Gale (excluding Windborne Material)

2.6.2.2 Lightning

Lightning is the second most significant weather-related cause for network faults, accounting for 3.76% of the faults recorded for the decade between 2010 and 2020. As illustrated in Figure 24, the monthly number of faults also demonstrates a seasonal trend, i.e., higher number of faults normally occurs in the months between May and September. This seasonal trend can be consistently observed over the 10-year period from 2010 to 2020. As mentioned previously, it is raised by the Met Office team that lightning forecast is difficult particularly for a 5-day window.

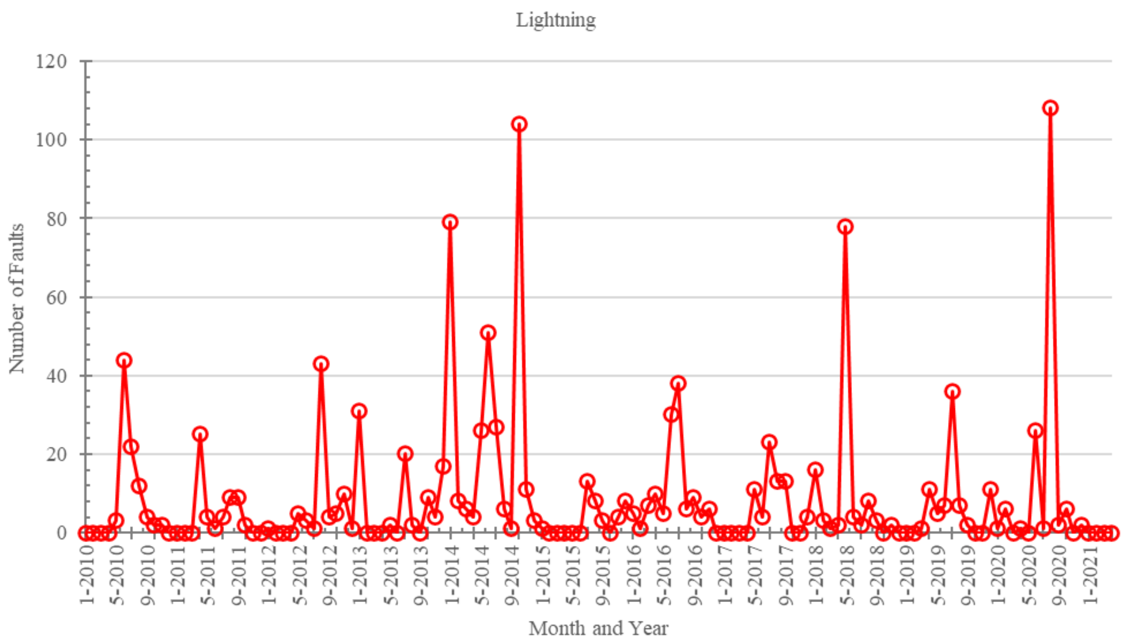


Figure 24: Monthly Number of Faults Caused by Lightning from 2010 to 2020

Figure 25 shows the statistics for the number of faults per month over the 10-year period from 2010 to 2020. It is found that 132 months had no more than 75 faults whilst 4 months had more than 75 but less than 150 faults. This implies that in general faults caused by lightning would not constitute an exceptional event.

Further statistics show that 75% of the months from 2010 to 2020 had less than 9 faults caused by Lightning, whereas 50% of the months had 3 or less such faults.

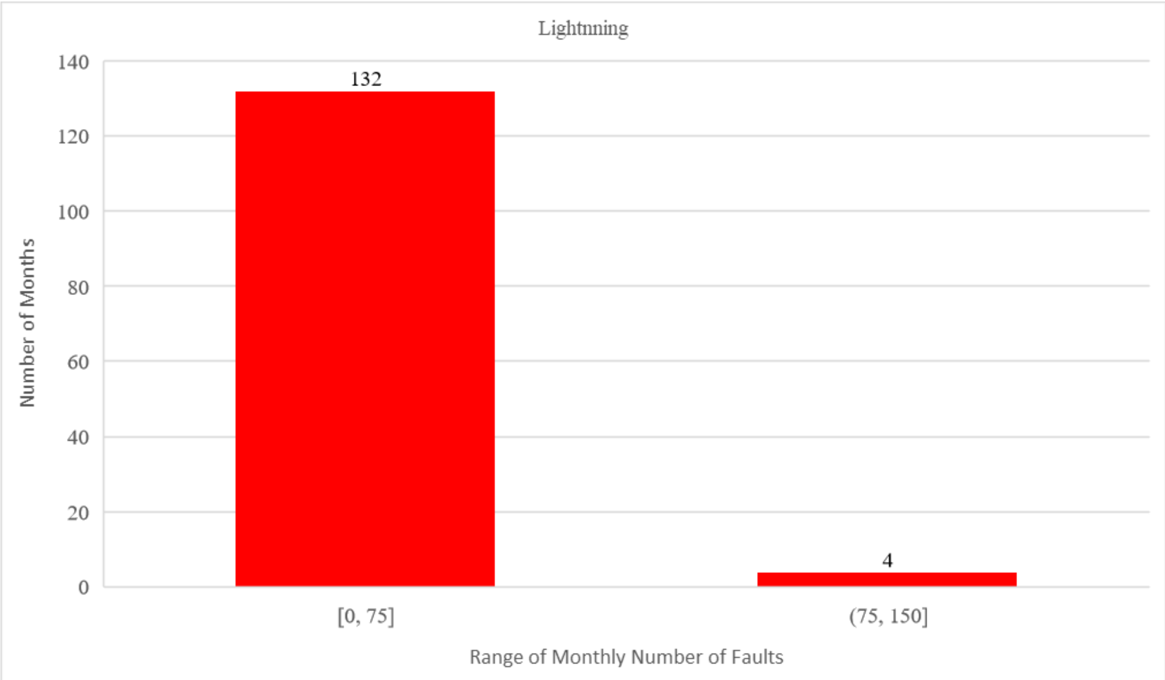


Figure 25: Statistics for Monthly Number of Faults Caused by Lightning

2.6.2.3 Snow, Sleet and Blizzard

The third weather-related cause for network faults is "Snow, Sleet and Blizzard", and the recorded number of faults per month is illustrated in Figure 26. The faults caused by Snow, Sleet and Blizzard are rare and in fact there were only 16 months with such faults recorded.

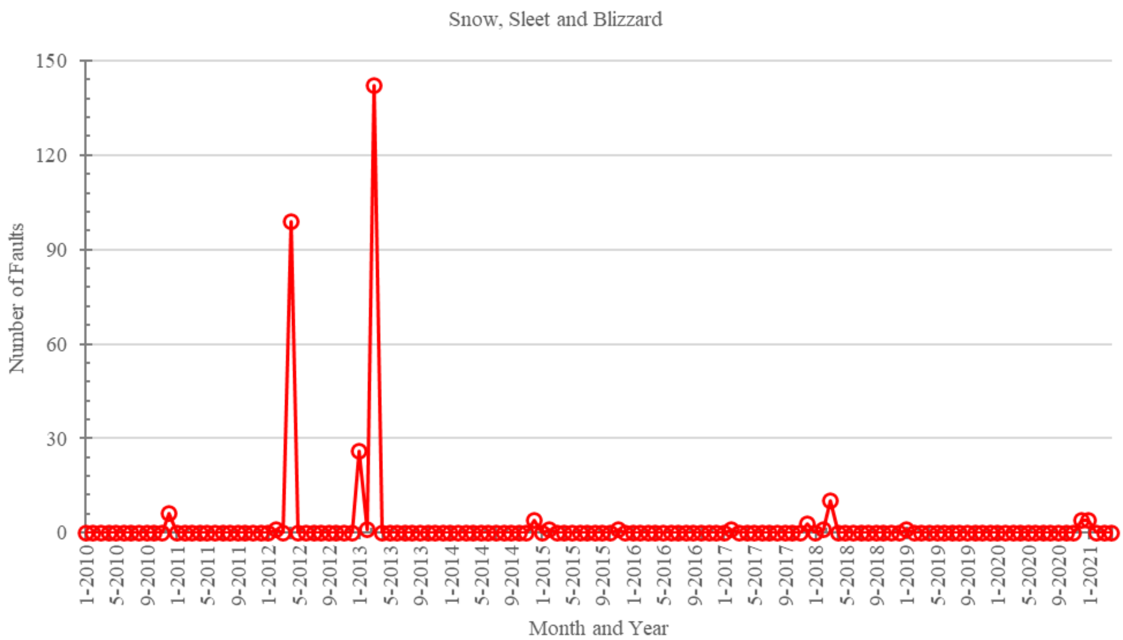


Figure 26: Monthly Number of Faults Caused by Snow, Sleet and Blizzard from 2010 to 2020

The monthly number of faults would not exceed 10 apart from April 2012, January and March 2013. This can be seen in Figure 27. This means that majority of faults caused by Snow, Sleet and Blizzard would not constitute an exceptional event; and the ability to forecast the number of faults that could be brought by Snow, Sleet and Blizzard should be able to improve network performance. Nonetheless, this also means that the forecast for this category of faults should be more focused on specific storms in order to maximise the benefit provided to the Control Room. Additionally, the location and magnitude of impact would be particularly important.

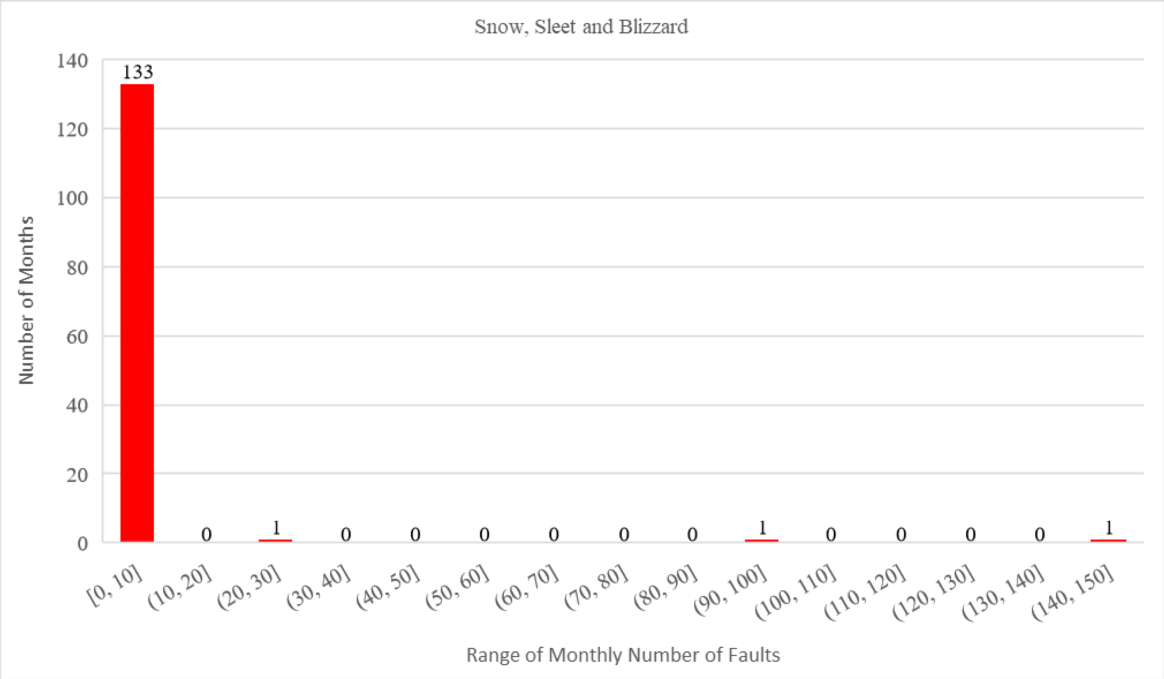


Figure 27: Statistics for Monthly Number of Faults Caused by Snow, Sleet and Blizzard

2.6.2.4 Windborne Material

Windborne Material is the fourth weather-related "Main Cause" for the recorded faults in the jurisdiction of SPM, totalling 245 faults over the 10-year period. Figure 28 shows that the number of faults per month varies but remains between 1 and 4 faults per month in the 10-year period, i.e., less seasonal variation.

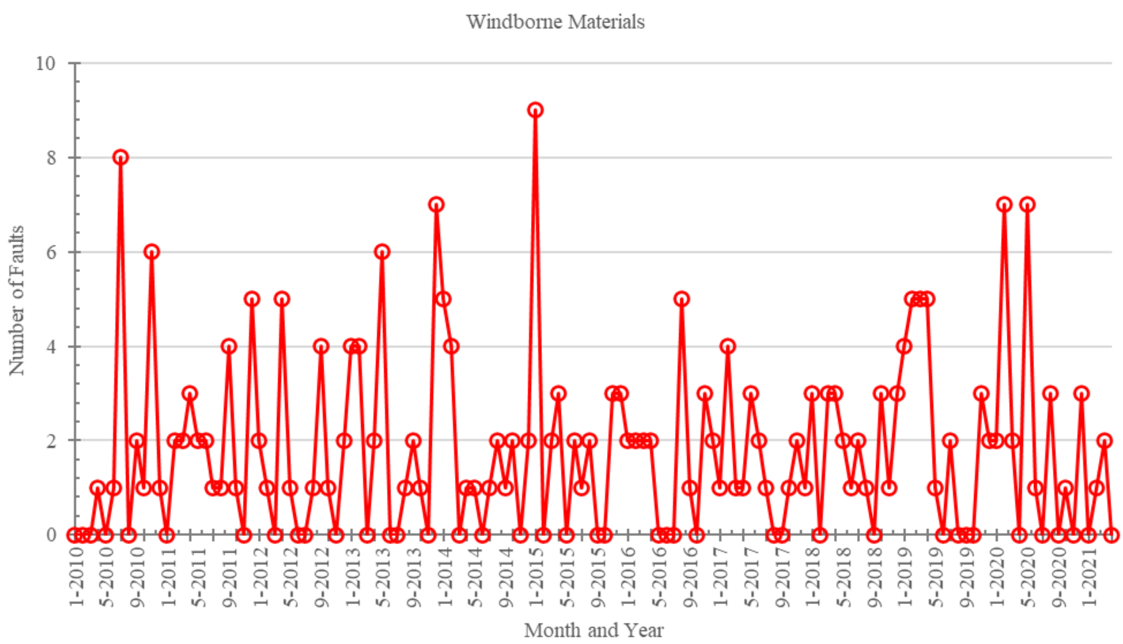


Figure 28: Monthly Number of Faults Caused by Windborne Material from 2010 to 2020

The distribution shown in Figure 29 adopts a bin size of 2 faults per month for the histogram, due to the low number of faults per month. According to Figure 29, 101 months had 2 or less faults caused by Windborne Material, and the highest number of faults per month is 9. Further, it is worth investigating how faults caused by Windborne Material correlate to those caused by Wind and Gale (excluding Windborne Material).

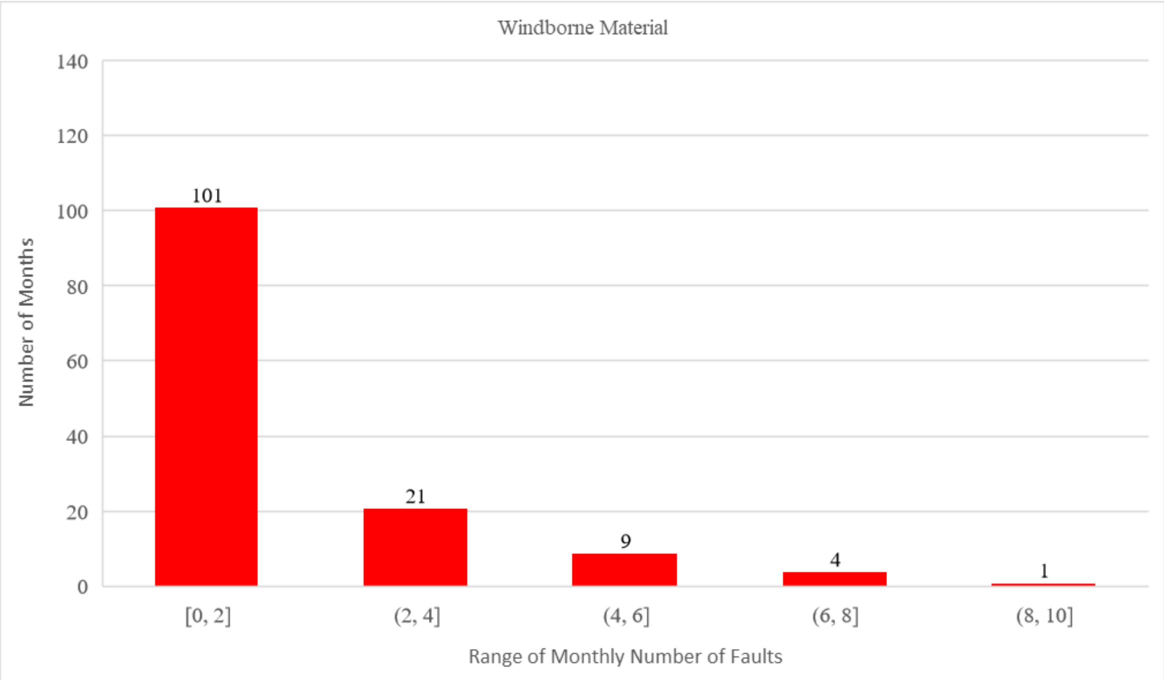


Figure 29: Statistics for Monthly Number of Faults Caused by Windborne Material

2.6.3 Geographical Distribution of Top Weather-related Faults

Prior to analysing the four top weather-related faults, the numbers of all the faults recorded in each of the six SPM districts are shown in Figure 30. The district of North Wales has over 10,000 faults, whilst the district of Dee Valley and Oswestry, and Mid Cheshire each has over 5000 faults recorded; these three districts accounts for over 65% of the total faults recorded for SPM. The district of Merseyside, Wirral, and Mid Wales each has 3,500 faults, or thereabouts.

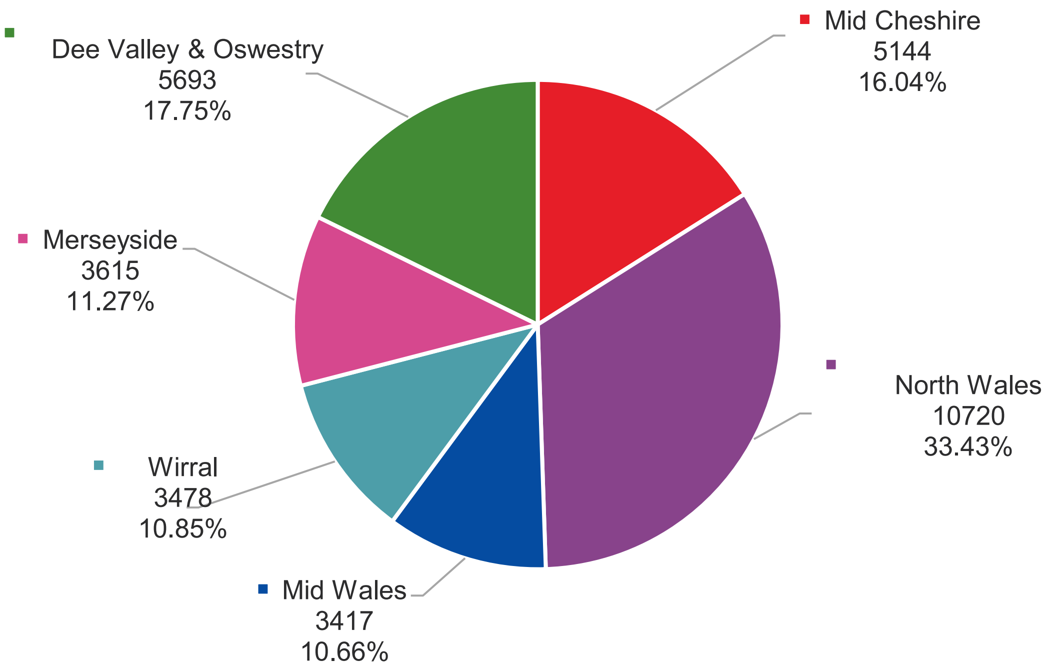


Figure 30: Number of Faults Recorded in Each of the SPM Districts

2.6.3.1 Wind and Gale (excluding Windborne Material)

According to Figure 31, the district of North Wales has the highest number of faults caused by Wind and Gale (excluding Windborne Material), followed by the district of Dee Valley and Oswestry, and Mid Wales. On the other hand, the district of Merseyside has the least number of such faults. The geographical distribution is mostly in line with that shown in Figure 30, i.e., these three districts account for over 70% of the faults caused by Wind and Gale (excluding Windborne Material).

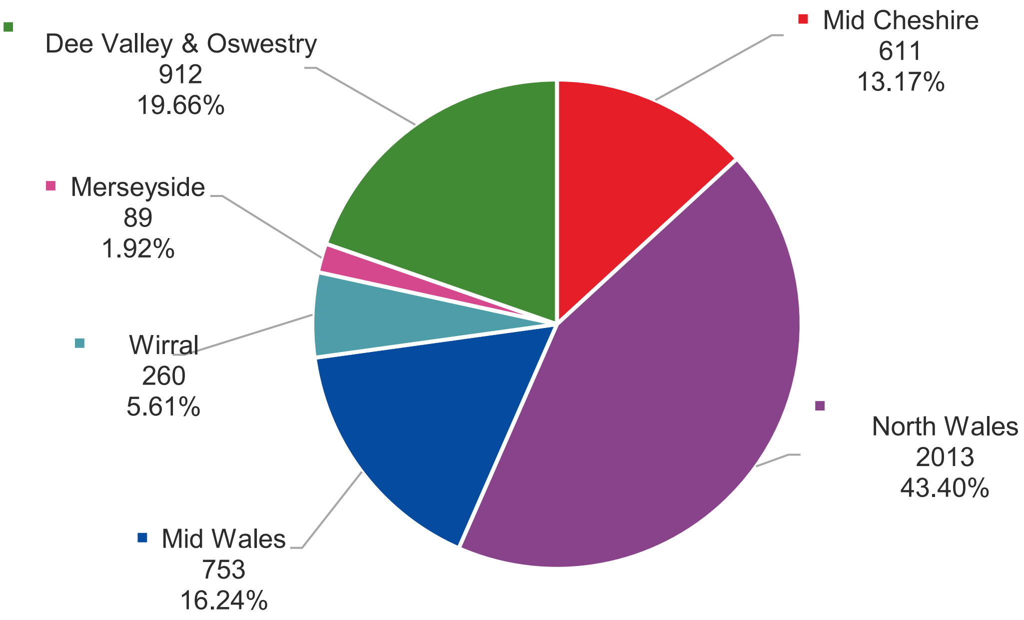


Figure 31: Number of Faults Caused by Wind and Gale (excluding Windborne Material) in Each of the SPM Districts

2.6.3.2 Lightning

Figure 13 shows the distribution of faults caused by Lightning across the six SPM districts. Half of the faults caused by Lightning are in the district of North Wales. The districts of Dee Valley and Oswestry, and Mid Wales account for another 32% of the total faults.

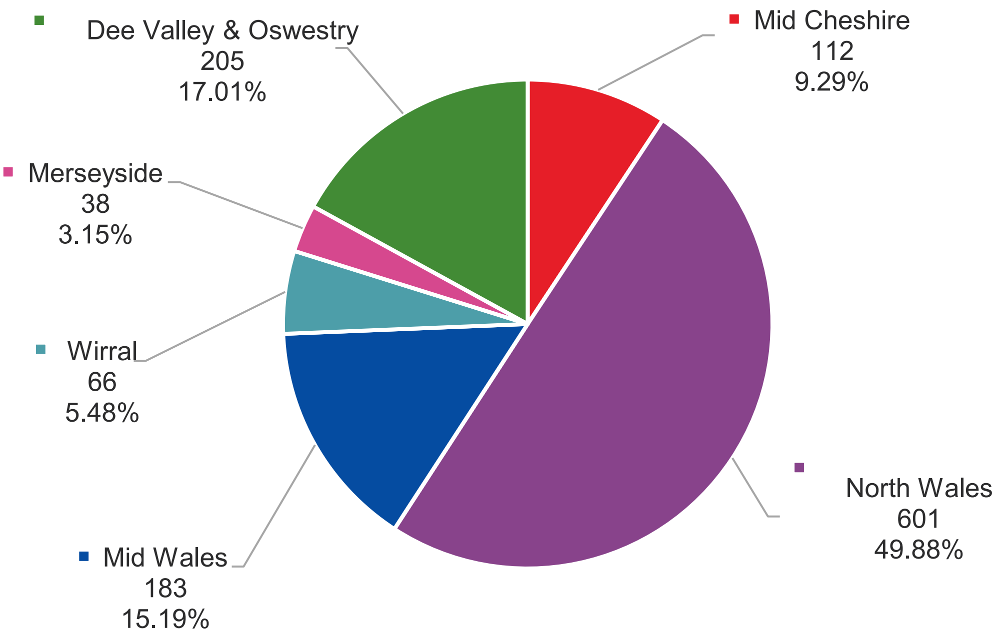


Figure 32: Number of Faults Caused by Lightning in Each of the SPD Districts

2.6.3.3 *Snow, Sleet and Blizzard*

The district of North Wales has 132 faults or 43.28% of the faults caused by Snow, Sleet and Blizzard, which is followed by the district of Dee Valley and Oswestry having 97 such faults, as shown in Figure 33.

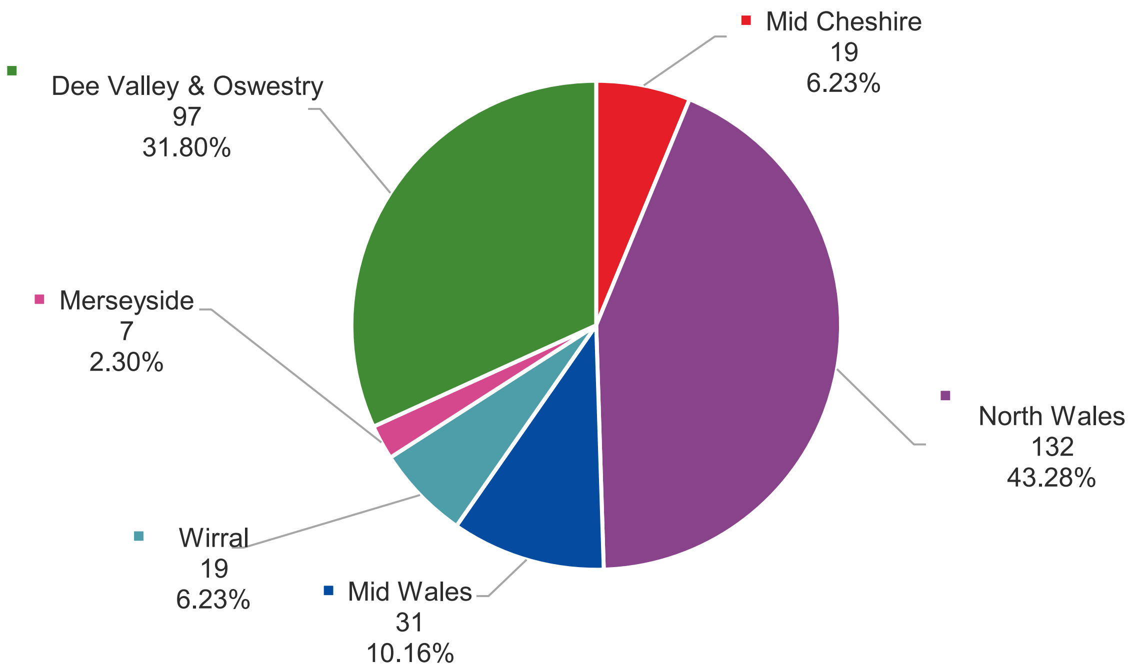


Figure 33: Number of Faults Caused by Snow, Sleet and Blizzard in Each of the SPD Districts

2.6.3.4 *Windborne Material*

The faults caused by Windborne Material distributed more evenly across the six SPM districts, whereas the district of North Wales accounts for close to 30% of the faults caused by Windborne Material, followed by the districts of Dee Valley and Oswestry, and Mid Cheshire.

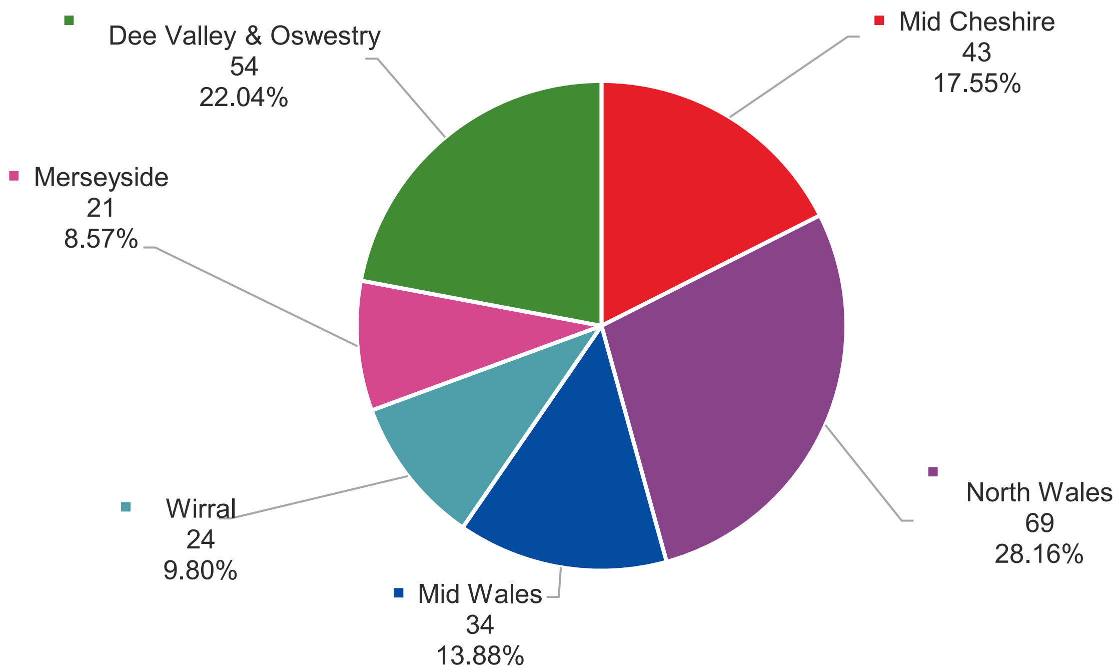


Figure 34: Number of Faults Caused by Windborne Material in Each of the SPD Districts

2.6.4 Impact of Top Weather-related Faults

This subsection presents the impact of the four top weather-related faults based on the longest duration, Customer Interruptions (CI), and Customer Hours Lost (CHL) provided in the historical network fault data. In order to analyse their impact, the box and whisker chart (also known as boxplot) is used, which is able to illustrate the distribution of the recorded data.

As mentioned previously, the bottom and top boundaries of the box represent the first and third quartiles, respectively; and the line in the middle of the box is the median value. The lower whisker corresponds to the minimum value, and the higher whisker is the boundary for outliers, which are represented by dots in the figures. The cross mark is the average value of the relevant data.

2.6.4.1 Wind and Gale (excluding Windborne Material)

The distribution of Duration Longest, CI and CHL is presented in Figure 35. Due to the long tail distribution of these metrics (i.e., the number of outliers is significant), the outliers shown in Figure 35 do not included all the outliers. This long tail distribution also leads to the average values of these metrics being skewed, compared with the corresponding median values. The maximum values of these metrics are as follows.

- Maximum of Duration Longest is 103.9 hours
- Maximum CI is 8,501
- Maximum CHL is 43,957.4

According to Figure 35, 75% of the Duration Longest is less than 10 hours and 50% of the Duration Longest is less than 4.5 hours, whereas, the Duration Longest between 2 hours and 10 hours accounts for half of the faults recorded. Three quarters of the faults resulted in a CI less than 202 and half of them had a CI less than 50; whilst three quarters of the faults resulted in a CHL less than 628 and half of them had a CHL less than 169. On the other hand, half of the faults show a range of CI between 12 and 202, and a range of CHL between 38 and 628.

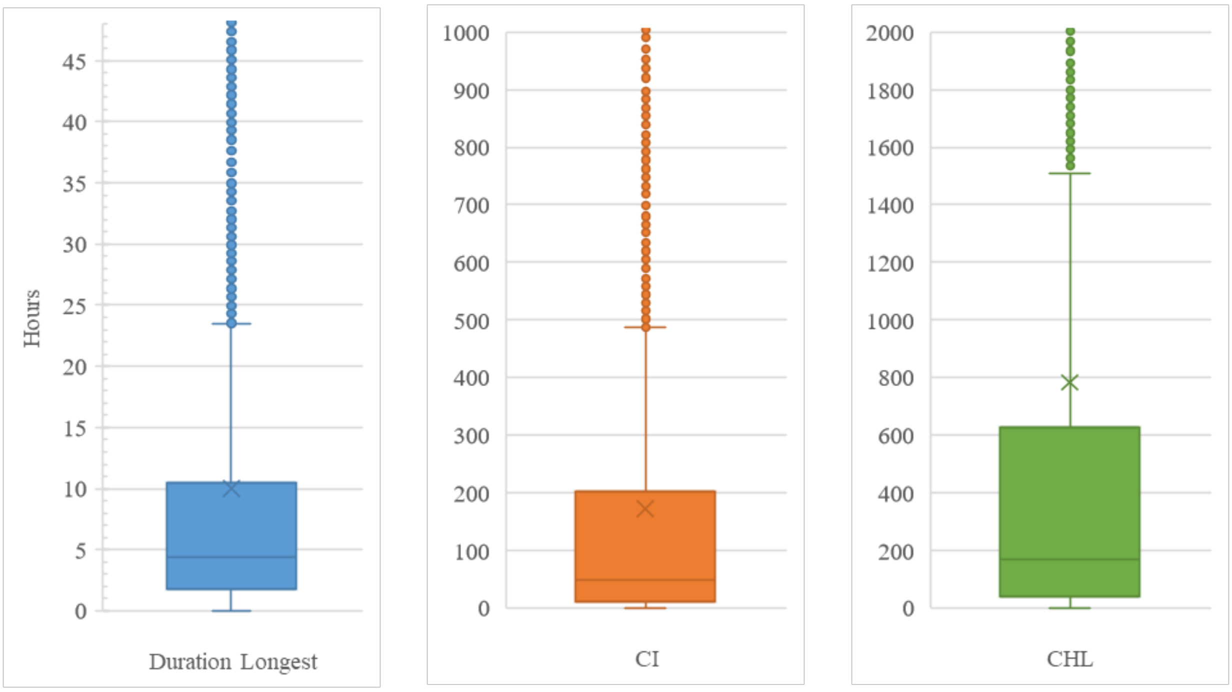


Figure 35: Boxplot for Duration Longest, CI and CHL of the Faults Caused by Wind and Gale (excl. Windborne Material)

2.6.4.2 Lightning

The magnitude of Lightning faults' impact is demonstrated in Figure 36. All the outliers are included for the Duration Longest, however only part of the outliers is shown in Figure 36 due to the long tail distributions. The maximum values recorded for these metrics are as follows.

- Maximum of Duration Longest is 42.2 hours
- Maximum CI is 9,058
- Maximum CHL is 12,439.1

According to Figure 36, 75% of the Duration Longest is less than 4.5 hours and 50% of the Duration Longest is less than 2 hours, whereas, the Duration Longest between 0.88 and 4.5 hours accounts for half of the faults recorded. Three quarters of the faults resulted in a CI less than 128 and half of them had a CI less than 24; whilst three quarters of the faults resulted in a CHL less than 132 and half of them had a CHL less than 41. On the other hand, half of the faults show a range of CI between 5 and 128, and a range of CHL between 8 and 132.

The faults caused by Lightning would have less impact from the Duration Longest and CHL's points of view, i.e., the faults were cleared and restored faster. However, the number of customers, as represented by CI, is still comparable to that of the faults caused by Wind and Gale (excluding Windborne Material).

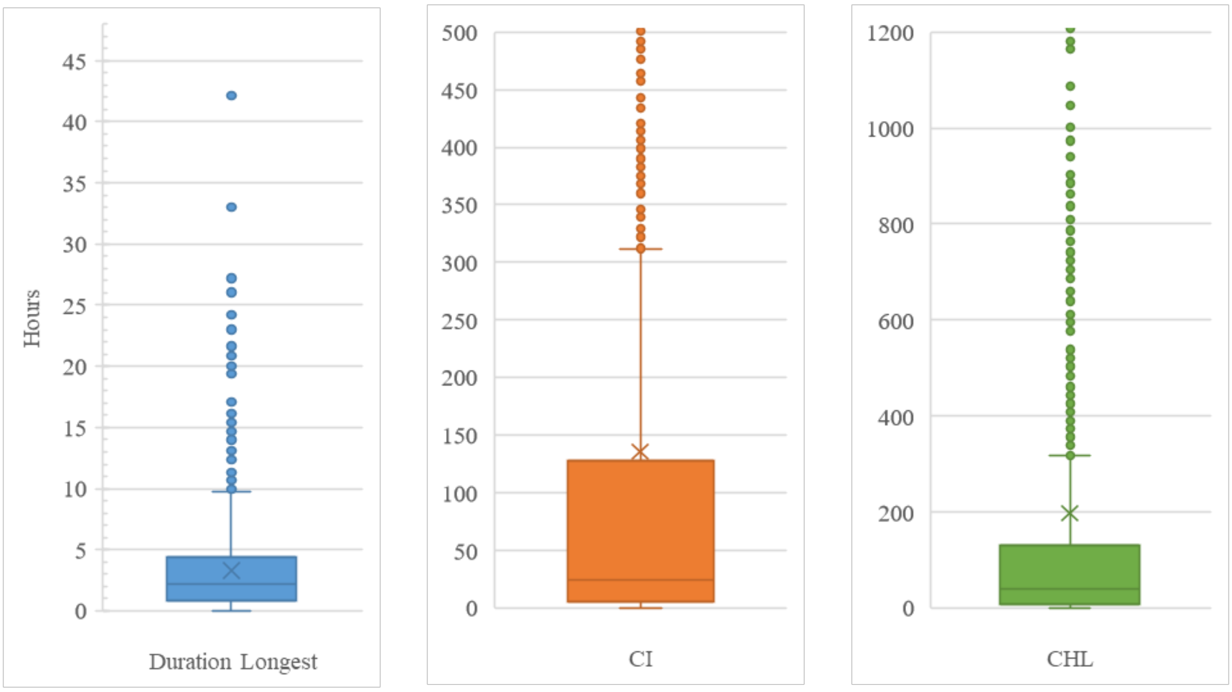


Figure 36: Boxplot for Duration Longest, CI and CHL of the Faults Caused by Lightning

2.6.4.3 Snow, Sleet and Blizzard

The distribution of Duration Longest, CI and CHL is presented in Figure 37. The long tail distribution of these metrics (i.e., the number of outliers is significant) lead to the outliers shown in Figure 37 not including all the outliers. As mentioned previously, this long tail distribution skews the average values of these metrics, compared with the corresponding median values. The maximum values of these metrics are as follows.

- Maximum of Duration Longest is 58.6 hours
- Maximum CI is 3,068
- Maximum CHL is 16,581.5

According to Figure 37, 75% of the Duration Longest is less than 13 hours and 50% of the Duration Longest is less than 6 hours, whereas, the Duration Longest between 3 hour and 13 hours accounts for half of the faults recorded. Three quarters of the faults resulted in a CI less than 199 and half of them had a CI less than 65; whilst three quarters of the faults resulted in a CHL less than 838 and half of them had a CHL less than 220. On the other hand, half of the faults show a range of CI between 17 and 199, and a range of CHL between 53 and 838.

It is observed that the Duration Longest and CHL is much higher than those of the faults caused by either Wind and Gale (excluding Windborne Material) or Lightning, whilst the CI is much lower.

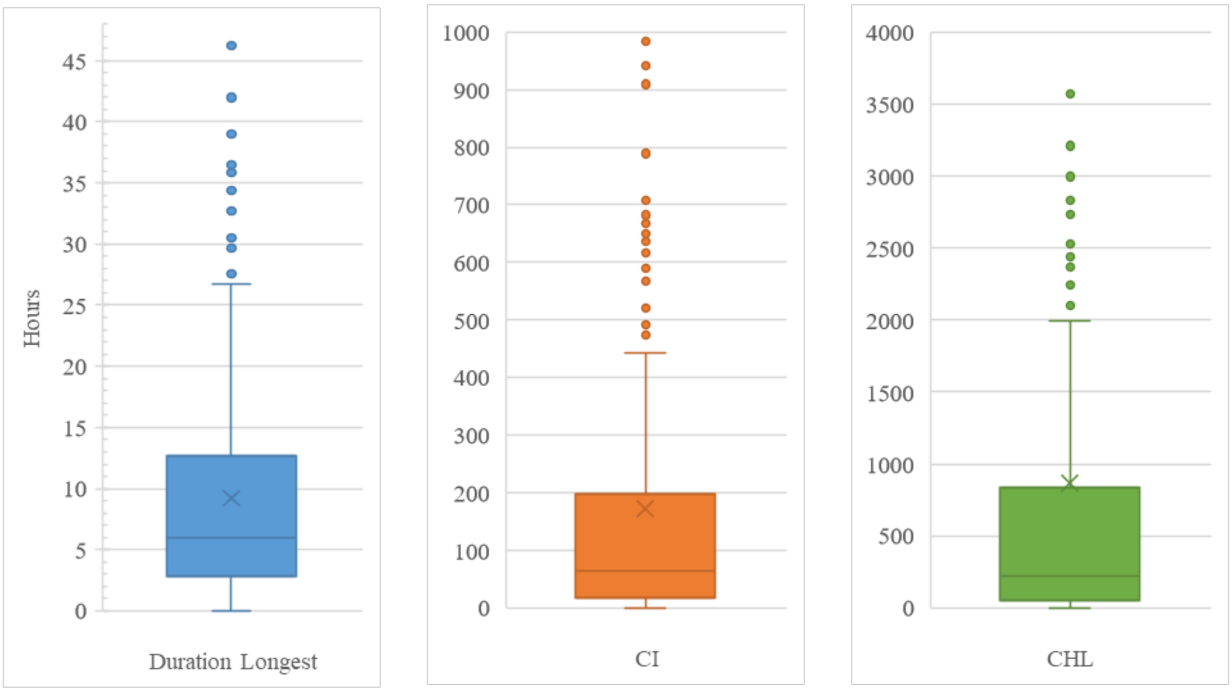


Figure 37: Boxplot for Duration Longest, CI and CHL of the Faults Caused by Snow, Sleet and Blizzard

2.6.4.4 Windborne Material

Similarly, Figure 38 does not show all the outliers for CI due to a long tail distribution, whilst it presented the whole distribution of the Duration Longest and CHL including all the recoded data. The maximum values of these metrics are as follows.

- Maximum of Duration Longest is 18.9 hours
- Maximum CI is 1,777
- Maximum CHL is 1,915.6

According to Figure 38, 75% of the Duration Longest is less than 3 hours and 50% of the Duration Longest is less than 2 hours, whereas, the Duration Longest between 0.7 and 3 hours accounts for half of the faults recorded. Three quarters of the faults resulted in a CI less than 140 and half of them had a CI less than 25; whilst three quarters of the faults resulted in a CHL less than 192 and half of them had a CHL less than 44. On the other hand, half of the faults show a range of CI between 5 and 140, and a range of CHL between 6 and 192.

The faults caused by Windborne Material had a much less significant impact with regard to their durations, which are approximately a third of the durations of the faults caused by Wind and Gale (excluding Windborne Material). However, from the CI's point of view, half of the faults either caused by Windborne Material or by Wind and Gale (excluding Windborne Material) would lead to a CI in the range between 5 and 140.

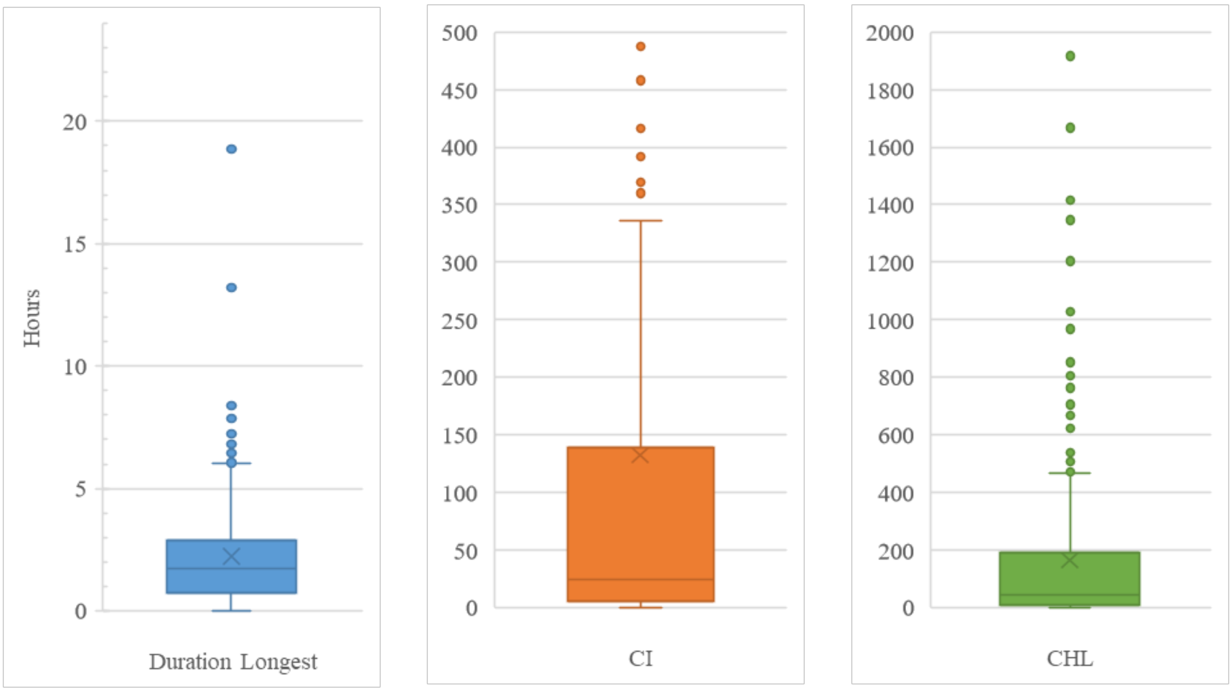


Figure 38: Boxplot for Duration Longest, CI and CHL of the Faults Caused by Windborne Material

3. Data Alignment

Historical network fault data need to be aligned with historical weather data both geographically and chronologically in order to establish their correlation. This will be the basis for transferring weather forecast to fault forecast. This section discusses the options for aligning the fault data with the weather data, and the recommendations for the Discovery Phase. A couple of issues found in the data are also noted in this section, which will be further investigated in future phases of the project.

3.1 Geographical Alignment

As listed in Section 2.3.1, there are three pieces of direct information to identify the geographical locations of the faults, which are as follows.

- District
- Location
- Grid Reference X and Grid Reference Y

“District” refers to the operational regions defined for SPD and SPM, whilst “Location” includes a brief description of the location of the relevant faults. “Grid Reference X” and “Grid Reference Y” are the coordinates recorded for these faults. As noted in Section 2.3.2,

Additionally, there are another four pieces of information which may also be used to help locate the faults. The following information is in essence asset-based and will require processing to find the relevant geographical locations. The asset-based information intrinsically will provide high granularity.

- Circuit Code
- HV Circuit Name
- Circuit Name (Incd.)
- HV Component

Furthermore, it has been raised that substation supply areas may be another option to mark the locations of faults. This provides a granularity higher than that provided by “District” but much less than the coordinates or the asset-based information.

3.1.1 Option for the Discovery Phase

During the Preparation Phase, multiple discussions were held with the Control Room team. It is noted that the granularity of how faults are grouped may have an impact on both the practicality for the fault forecast and the accuracy that may be achieved. For example, one of the difficulties is whether the recorded data would be sufficient to support forecast at that level of geographical granularity. Additionally, factors other than those related to weather (e.g., asset health condition) would need to be considered as input to the forecast, and this introduces uncertainty and complexity, which may not bring the desired benefit.

According to the discussions with the Control Room team, the project team agreed that the fault forecast would be based on the districts defined by SPEN for the Discovery Phase, in order to provide a proof of concept. The more granular option of using the substation supply areas may be investigated in future phases of the project. Additionally, forecast based on the operational districts may be most beneficial from the perspective of mobilising recourses.

3.1.2 Discrepancy in the Geographical Information

The geographical information included in the historical network fault data is not complete as noted in Section 2.3.2. Additionally, the GIS shapefiles are provided to support the extraction of historical weather data. Figure 39 and Figure 40 show the GIS shapefiles covering the districts defined for SPD and SPM, respectively.

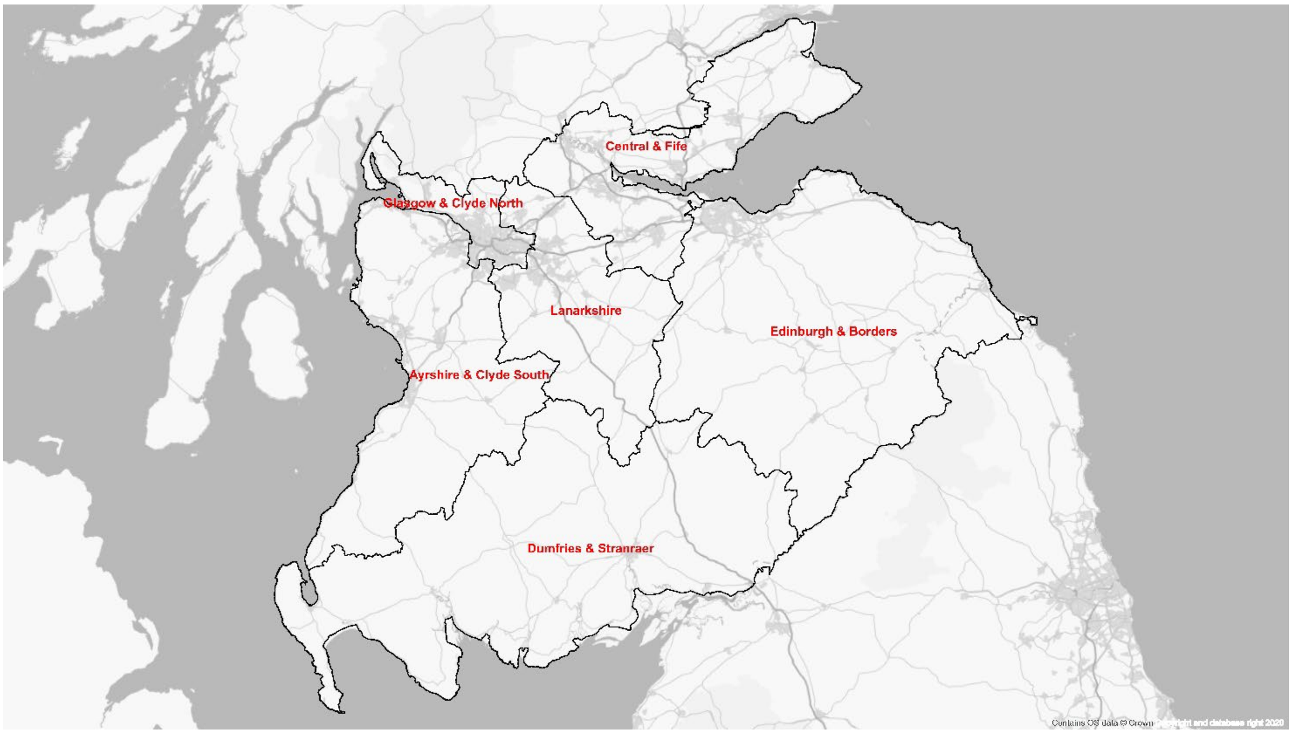


Figure 39: SPD Districts in the GIS Shapefile

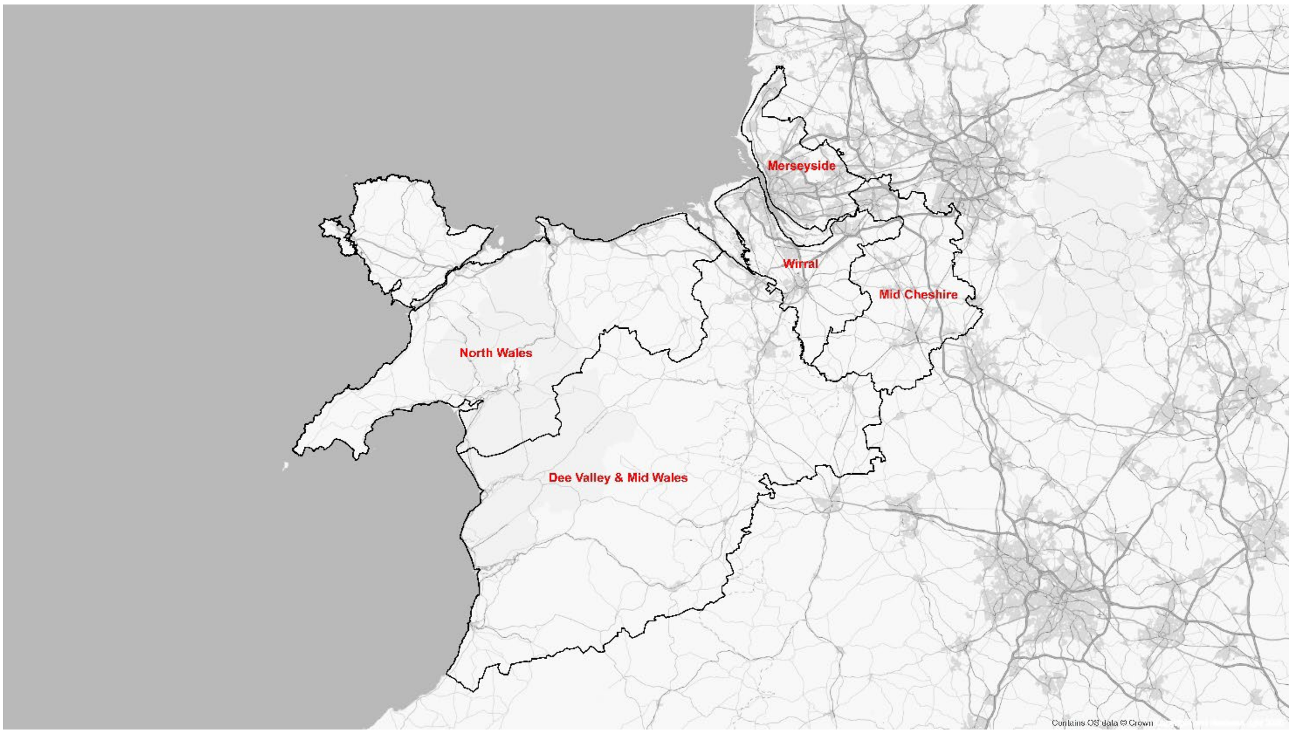


Figure 40: SPM Districts in the GIS Shapefile

Comparing to the districts included in the historical network fault data, the discrepancies are found in the GIS shapefiles. More specifically, there are 7 districts noted in the fault data of SPD, whereas in the GIS shapefile the districts of Edinburgh (and Lothian) and Borders are merged, as seen in Figure 39. Similarly, the fault data of SPM include two separate districts of Mid Wales and Dee Valley (and Oswestry); however, the GIS shapefile shows them as one district. The above discrepancy may not have any substantial impact on the Discovery Phase. However, these should be addressed in the next phase of the project.

3.2 Chronological Alignment

As listed in Section 2.3.1, the “Date/Time of Incident” and “Reporting Year” included in the historical network fault data are time related, whereas the former one is a timestamp accurate to minutes. As discussed in Section 2.3.3, inconsistency is discovered between “Date/Time of Incident” and “Reporting Year”. However, the timestamp (i.e., “Date/Time of Incident”) is used as the basis to align the fault data with the weather data.

3.2.1 Option for the Discovery Phase

The key aspect of aligning the fault data with the weather data is to decide the period of time for grouping the number of faults, which is the focus of the fault forecast. This period of time could be a quarter of an hour, half an hour, hourly, half day, daily, and so forth.

During the Preparation Phase, it was agreed with the UOG team that the preferred option would be to group the number of faults on an hourly basis, which then could also be grouped into other longer period of time should that be needed.

4. Review of Software and Modelling Techniques

Software and modelling techniques play a critical role in the delivery of research and development projects. It is particularly important for the Discovery Phase of the Predict4Resilience project in order to minimise any abortive work. This section reviews some typical software and modelling techniques with in the context of power system analysis, and discusses their suitability for the Predict4Resilience project.

4.1 Modelling Techniques

Multiple modelling techniques of power system analysis could be useful to support the Predict4Resilience project. These may include but not limited to:

- Monte-Carlo simulation: it can simulate the impact of weather events on the distribution network.
- Power flow calculation: it can calculate static power flows within a distribution network.
- Optimal power flow: it can calculate any re-dispatch or network re-configuration following a failure.
- Fault level calculation: it can calculate fault currents in the network follow a fault.

These modelling techniques are not mutually exclusive and can be combined to provide a specific assessment as needed. Primarily, they can be used to provide a theoretical assessment of any impact on the network and inform any benefit that may be brought by a change to the operational practices, such as the use of the Weather Fault Tool.

4.2 Software

There are many pieces of commercially available software in the market that can facilitate the above modelling techniques, including but not limited to these as follows.

- MATLAB or any Python programming environment, etc.
- DigSILENT
- IPSA or ERACS or ETAP, etc.

Bespoke scripts can be developed in MATLAB or Python to exactly suit the project's need. DigSILENT is a powerful and comprehensive graphic-based electrical network modelling software, whilst these included in the third bullet point have less functionalities but can delivered specific calculations as required.

4.3 Project Preference

Detailed discussions were held by the project team with the Control Room team during the Preparation Phase. The objective of the Predict4Resilience project is to transfer weather forecast to fault forecast, more specifically the forecast of number of faults within a time window for a geographical region. This means that a data-driven method will be applied by using a type of machine learning technology and training the forecast engine with historical data.

Therefore, the modelling techniques and software mentioned in Sections 4.1 and 4.2 will not be relevant to the core of this project, and they will not be needed for the Discovery Phase. Nonetheless, Monte-Carlo based techniques may be useful in later phases of the project should there be a need to evaluate in detail the benefits that could be brought by deploying this Weather Fault Tool.

5. Conclusion

In the Preparation Phase, a detailed review of the historical network fault data has been performed, and the project team has engaged with the Control Room team multiple times in order to shape the Discovery Phase. This section summarises the findings of the data review and presents the preferred options for the Discovery Phase.

5.1 Data Review Findings

The key findings of the data reviews include:

- Historical network fault data are of relatively good quality.
- Geographical locations at district level and timestamps are generally recorded in a consistent manner. However, two thirds of the Grid References X and Y are blanket, whilst some of them may contain errors.
- The top "Main Cause" for faults is "Deterioration due to Ageing or Wear (excluding corrosion)", whilst "Cause Unknown" has the second or third highest number of recorded faults.
- The top three weather-related "Main Cause" for faults are as follows:
  - Wind and Gale (excluding Windborne Material)
  - Lightning
  - Snow, Sleet and Blizzard
- The top three SPD districts with high number of weather-related faults are as follows:
  - Dumfries
  - Ayrshire and Clyde South
  - Central and Fife
- The top three SPM districts with high number of weather-related faults are as follows:
  - North Wales
  - Dee Valley and Oswestry
  - Mid Wales
- Three quarters of the top three weather-related faults have comparable magnitude of impact on the network based on longest duration, CI, and CHL.

5.2 Interface between Fault Data and Weather Data

The historical network fault data have been grouped by their "Main Cause", and subsequently the number of faults has been counted on an hourly and by district basis. These numbers of faults are used to establish the statistical association with the relevant weather data, in order to provide a proof of concept for the Discovery Phase.

5.3 Preference for Modelling Techniques and Software

Although multiple modelling techniques and software for power system analysis are available, the Predict4Resilience project's objective dictates that a data-driven method based on machine learning should be the core modelling technique. There will not be much need for power system analysis in the context of the fault forecast.

5.4 Potential Next Steps for Alpha Phase

As mentioned in earlier sections of the report, there may be a few points worth further investigation during the next phase of this project. However, these do not affect the objective of the Discovery Phase for providing a proof of concept. The two key steps may include:

- To automate historical network fault data processing
- To improve the geographical granularity of the historical network fault data in accordance with the need of the forecast model.