Climate Resilience Assessment to Inform Asset Upgrade Strategy Development for Infrastructure Systems

A thesis submitted in partial fulfilment for the degree of Doctor of Philosophy

in the

Department of Civil and Structural Engineering

Qianqian Li

July 2023

Declaration of Authorship

The work this thesis is the author's own work unless reference has been made to the work of others.

I, the author, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means. This work has not previously been presented for an award at this, or any other, university.

Signature: Qiangian Li

Acknowledgement of Collaborative Work within the Thesis

The work in this thesis has resulted in the following peer-reviewed publications:

1. Li, Q., Punzo, G., Robson, C., Arbabi, H. and Mayfield, M., 2022. A Systematic Approach to Climate Resilience Assessment of Infrastructure Networks, IEEE Systems Journal. (Accepted)

2. Li, Q., Punzo, G., Arbabi, H, 2023. Unified approach to optimal resource allocation for climate resilience of transport networks. (In preparation for submission to the Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences)

In publications, I, the candidate, planned and wrote the manuscript, co-author reviewed and provided comments on the manuscript.

Abstract

Current infrastructure systems may not have the ability to withstand the future climate, where unprecedented extreme weather events with larger magnitudes and higher frequencies, may happen in new geographical locations, at new times of the year. Considering the centrality of infrastructure systems in society, it is necessary to understand how future climate and the consequential patterns of weather extremes impact infrastructure systems and develop tools to guide investment decisions to minimise disruption in the future. This PhD project addressed the following two research questions. 1) How can the magnitude of future infrastructure service disruption due to climate change be estimated accounting for corresponding uncertainties? 2) How should resources be allocated in the most efficient manner to upgrade infrastructure assets so that the forthcoming service disruptions can be minimised?

To answer the first research question, a systemic approach to assess the resilience of infrastructure systems to climate change is formulated. The proposed approach advances the current resilience assessment approaches where failure scenarios are primarily simulated as random or targeted removal of system components, or simulated with manipulations of present-day weather statistics. A case study is then carried out to quantify the resilience of Great Britain's railway passenger transport system to high-temperature-related track buckling under the Representative Concentration Pathway 8.5 (RCP8.5) climate change scenario. A 95-year horizon on the resilience of the railway system is drawn. The results reveal the non-linear responses of the railway system to the increasing temperature and show that models considering random asset failures overestimate the system's resilience.

Two complementary approaches were identified as relevant to the second research question - the node-edge ranking approach in the network science community and the mathematical optimisation approach for resource allocation problems in the field of operational research. From the perspective of node-edge ranking, metrics that capture the topological structure of the network, supply-demand features, and future weather failure patterns are developed to guide resource allocation. The resource allocation problem is formulated as an optimisation problem that takes the future weather pattern as model inputs and aims to maximise the climate resilience of the system. The problem is then analytically solved in a simple model. The analytic solution from the simple model converges to the same expression as one of the developed metrics. The two streams of work implementing the two approaches meet each other. Based on the observed consistency in solutions obtained with the two approaches, a ranking-based resource allocation

strategy is developed and applied to a core freight network of Great Britain. The results suggest that the newly proposed approach can effectively tackle the resource allocation problem for large-scale infrastructure systems with low computational costs.

The proposed framework and findings from the case studies advance the existing understanding of the climate resilience of infrastructure systems and strategies on resource utilisation for climate change adaptation. The implementation of such strategies is imperative in avoiding the worst consequences of climate change.

Contents

Declaration of Authorship iii

Acknowledgement of Collaborative Work within the Thesis v

Abstract vii

Table of Contents ix

List of Figures xiii

List of Acronym xviii

1 Introduction 1

1.1 Background 1

1.2 Research Aim and Objectives 3

1.3 Research Contribution 4

1.4 Thesis Structure 5

2 Climate Change and Infrastructure Systems 7

2.1 Chapter Introduction 7

2.2 Climate Change and Infrastructure Systems 7

2.2.1 Climate Hazards to Infrastructure Systems 7

2.2.2 Impact of Climate Change on Infrastructure Systems 9

2.2.3 Climate Change Impact Assessment 12

2.2.4 Brief Summary 13

2.3 Resilience as A Tool for Climate Change Impact Assessment 13
2.3.1 The Concept of Resilience 14
2.3.2 State of the Art: Resilience Assessment Frameworks 16
2.3.3 A Brief Summary 22
2.4 Resource Allocation Strategies towards Climate Resilience 23
2.4.1 Network Metrics as Importance Measures 23
2.4.2 Optimisation Problems for Networked Systems 28
2.4.3 Brief Summary 34
2.5 Research Gaps 35
2.5.1 A Systematic Climate Resilience Assessment Framework 35
2.5.2 A Resource Allocation Strategy towards Climate Resilience 36
2.6 Chapter Conclusion 36

3 Climate Resilience Assessment Framework 37

3.1 Chapter Introduction 37
3.2 A Systematic Climate Resilience Assessment Framework 38
3.2.1 Network Representation of Infrastructure Systems 38
3.2.2 Failure Scenario Generation 40
3.2.3 Probabilistic Resilience Measure 41
3.3 Resilience of The UK Rail Passenger Network to High Temperatures 43
3.3.1 Model Inputs and Assumptions 43
3.3.2 Simulation 50
3.3.3 The 95-year Trend 51
3.3.4 Shock-disruption Relationships 52
3.3.5 Climate-based, Random and Targeted Failures 55
3.3.6 Case Study Discussion 57
3.4 Chapter Conclusion 61

4 Asset upgrade strategy towards climate resilience 63

4.1 Introduction 63
4.2 Formulation of the Resource Allocation Problem 64
4.2.1 The System Model 64
4.2.2 The Objective Function 65
4.2.3 The Resource Allocation Problem 68
4.3 Development of Ranking Strategies 70
4.3.1 OD Centrality 71
4.3.2 Service Degradation 72
4.3.3 Gradient 72
4.4 The Analytical Solution of A Test Model 73
4.4.1 The Objective Function 74
4.4.2 Direct Application of the Lagrange Multiplier Method 76
4.4.3 The Modified Objective Function 78
4.5 The Unified Approach 83
4.6 Chapter Conclusion 85

5 Case Studies For Asset Upgrade 87

5.1 Case Study 1 - A Simple Network Model 88
5.1.1 The Network Model and the Optimisation Problem 88
5.1.2 The Optimisation Problem 88
5.1.3 Solving the Optimisation Problem 90
5.1.4 Results 93
5.1.5 Discussion 95
5.2 Case Study 2 - Network with up to 6 Nodes 97
5.2.1 The Network Models 97
5.2.2 The Numerical Method v.s. the Metric-based Method 97
5.2.3 The Iterative Metric-based Resource Allocation Method 99
5.2.4 Discussion 101
5.3 Case Study 3 - UK Core Freight Rail Network 101

5.3.1 The Objective Function . . . 101
5.3.2 The Network Model . . . 102
5.3.3 The Climate Hazard and Climate Data . . . 104
5.3.4 Numerical Solve: Heuristic Algorithm . . . 105
5.3.5 The Iterative Metric-based Solve . . . 105
5.3.6 Results - Heuristic v.s. Iterative Metric-based . . . 108
5.3.7 Results - Optimising towards Extreme Days and Toward Typical Days . . 110
5.3.8 Results - Correlations with Constant Variables . . . 111
5.3.9 Discussion . . . 115
5.4 Chapter Conclusion . . . 118

6 Discussion **121**

6.1 Chapter Introduction . . . 121
6.2 Addressing Infrastructure Network Resilience to Climate Change . . . 122
6.3 Limitations of the Proposed Approach . . . 125
6.4 Challenges in Implementing the Framework . . . 127
6.5 Applicability to General Infrastructure Systems and Weather Hazards . . . 128
6.6 Chapter Conclusion . . . 129

7 Conclusion **131**

7.1 Main Contributions and Findings . . . 131
7.2 Recommendations for Future Works . . . 133

List of Figures

1.1 Four indicators of global climate change from Masson-Delmotte et al. (2021) . . 2

1.2 A simplified process on how the climate system impacts the infrastructure systems 2

1.3 Chapter Structure of this Thesis 5

2.1 Conceptual map of interacting weather phenomena impact on road infrastructure and transport. Recreated from Leviäkangas et al. (2011). 9

2.2 Schematic showing the effect on extreme temperatures (a) the mean temperature increases, (b) the variance increases, and (c) when both the mean and variance increase for a normal distribution of temperature (Houghton et al., 2001) 10

2.3 Probability ratio (PR) of exceeding the 99th (blue) and 99.9th (red) percentile of pre-industrial daily (**a**) temperature and (**b**) precipitation at a given warming level, averaged across land. The probability ratio is defined as the probability of exceeding a certain quantile with a given level of global warming over the probability of exceeding it during the pre-industrial control period (Fischer and Knutti, 2015). Figures from Fischer and Knutti (2015) 11

2.4 The conceptual definition of the measure of resilience proposed by Bruneau et al. (2003). The curve shows the variation of the quality of the infrastructure (Q(t)) from the time of the event happening ($t=t_0$) to the time of full recovery ($t=t_1$). The quality of infrastructure ranges from 0% (no services) to 100% (fully functional) and is measured as the proportion if infrastructure service delivered. Figure from Bruneau et al. (2003). 15

2.5 The network model of the supply chain in Nagurney et al. (2002), which consists of m manufactures (the top row of nodes), n retailers (the middle row), and o demand markets. Figure from Nagurney et al. (2002). 17

2.6 A simple example of two interdependent systems with four types of dependency. For example, the flow between N1 and N2 utilises asset A1 in System 1. Figure from Goldbeck et al. (2019). 18

2.7 Example fragility curves. (a) A fragility curve in the form of a step function, which can be used to represent brittle and abrupt failures. (b) Three s-shaped functions representing probabilistic failure modes. 19

2.8 The wind fragility curves of transmission lines and towers for the normal and robust networks by Panteli and Mancarella (2015b) 20

2.9 Conceptual resilience curve where the recovery phase is presented with uncertainty. Figure from (Aydin et al., 2018). 22

2.10 Concave and convex function 33

3.1 Overview of the probabilistic resilience assessment framework. Failure scenarios are sampled with weather profiles across the network. Service disruption is then calculated for each sampled failure scenario, forming a probability distribution. The resilience of infrastructure systems to future climate is assessed with a collection of such probability distributions. 39

3.2 A illustration for projecting the weather profile on the asset layer and converting the local weather parameters to probabilities of failure. 40

3.3 A illustration for the sampling process. 42

3.4 (a) Asset layer of the network model with 4024 nodes and 4524 edges. Edges weights indicate the amount of flows on edge with a minimum of 0.2 pixel. (b) Flow layer of the network model. This figure only plots 22,326 out of the total 2,282,270 OD pairs, which has more than 15 passenger trips and geographical path length greater than 30km. 44

3.5 Clustering of the summer days (01 May to 30 September) from 2006 to 2100 based on the tasmax of EURO-CORDEX output data EUR-11_CNRM-CERFACS-CNRM-CM5_rcp85_r1i1p1_MOHC-HadREM3-GA7-05_v2. 15 clusters are produced for each 5-year group through a K-means algorithm. The height of each bar is controlled by the number of days that fall into the cluster. The color of the bar is controlled by the average temperature of the synthetic centroid of the cluster. . 47

3.6 Histogram of the percentage difference between the amount of flow delivered at each time step between a full rerouting effort and the reduced rerouting effort. Twelve failure scenarios are included in the sample. The number of trips delivered with the two rerouting efforts is compared at every time step. 49

3.7 Illustrative resilience curve with calculated LOS as the area of the shaded 'triangles'. 51

3.8 Time series of the estimated annual total LOS for the 2006-2100 period. The solid black line shows the expectations of the aggregated distributions. The shaded area shows the 5% to 95% range across the aggregated distribution. The solid red line is obtained by smoothing the solid black line with a Savitzky–Golay filter (Savitzky and Golay). 52

3.9 Relationship between (a) The average of measured cumulative loss of service and the national average tasmax; (b) the expectation of the number of edge removed Ψ and the national average tasmax; (c) The average of measured cumulative loss of service and expectation of the number of edge removal, Ψ; (d) the expectation of the number of edge removed Ψ and the average of loss of service at the onset of disruption; (e) the average of cumulative loss of service and the onset loss of service. The national average tasmax is the average of the tasmax across the asset layer. 53

3.10 Mapping of the tasmax for three days with similar national average tasmax but different LOS. 54

3.11 Comparing the the initial onset loss of service between the climate-based, random and targeted failure scenarios. Upper - the percentage of OD journey interrupted at the onset of disruption without any rerouting or repairing effort against the number of nodes removed. The sample size is 250. The shaded area shows the 2.5% to 97.5% range. Markers are the means of each sample. Lower - p-values of the Mann-Whitney U test between samples from climatic and random failure scenarios vs. the number of nodes removed. Any p-value smaller than 0.0001 is replaced with 0.0001. The red horizontal line is p=0.01, the confidence level adopted. 55

3.12 Mapping of the tasmax for three days that the climate-based failure scenarios turn out to be (a) more disruptive, (b) less disruptive, and (c) as disruptive as the random failure scenarios. 56

3.13 Comparing the cumulative loss of service between the climate-based, random and targeted failure scenarios. Upper - the measured LOS against the number of edges removed. The sample size is 250. The shaded area shows the 2.5% to 97.5% range. Markers are the means of each sample. Lower - p-values of the Mann-Whitney U test between samples from climate-based and random failure scenarios against the number of nodes removed. Any p-value smaller than 0.0001 is replaced with 0.0001 for visualization purposes. The red horizontal line is p=0.01, the confidence level adopted. 57

3.14 The histograms of calculated sample size with power analysis. Left - the calculated required sample size for the normally distributed pairs. Right - the calculated required sample size for the gamma distributed pairs. 60

3.15 The difference between the percentage of journey interrupted obtained from N=1000 and the percentage of journeys interrupted with N=250, for the climate-based scenarios and the random removal scenarios. 61

4.1 The structure of Chapter 4 64

4.2 A simple network illustration for the bi-layer structure. 65

4.3 The network structure of the test model. The flow layer has 3 OD pairs. The asset layer is a network of 3 nodes and 3 edges. 74

5.1 The network structure of the testing model. The flow layer has 3 OD pairs. The asset layer is a network of 3 nodes and 3 edges. 88

5.2 Equation 5.7, the fragility function with $\omega_1 = 45, \mu_1 = 35, \sigma_1 = 2.5$. The x-axis is the value of $\Delta\mu_1$ and the y-axis is the probability of failure. 92

5.3 Histogram of percentage difference between the objective function with solutions obtained from the three different methods. The comparisons are made between the values of the objective function and calculated as percentage difference. . . . 94

5.4 (a) Strip-plot of the percentage difference between the objective function with solutions obtained from the numerical method and the metric-based method, grouped by m. The diamond markers are the averages of each group. (b) Scatter plot of the percentage difference between the objective function with solutions obtained from numerical method and the metric-based method verse the maximum value of $\Delta\mu_i$ 96

5.5 The graph atlas generated using codes from NetworkX (NetworkX, 2023) 98

5.6 Percentage differences between numerical and metric-based method. 99

5.7 a) Values of the objective function (percentage of service disruption) against the number of iterations used in the iterative metric-based method. The black line on top is the averaged initial service interruption without any upgrade (m=0). The dashed horizontal line are the objective functions minimised using SQP. The solid lines are the objective functions achieved by running the iterative metric-based method. b) Time taken (seconds). 100

5.8 (a): The key freight corridors identified by Network Rail Network Rail (2017b). (b): The asset layer constructed for this study (red) embedded in all railway track lines across the country that is managed by Network Rail (grey). The original shapefile is provided by Network Rail. 102

5.9 Cluster of stations near Southampton 103

5.10 (a) The average of the objective functions from all 25 days versus the values of *m_stepsize*. b): The average of the simulation times for all 25 days versus thee values of the *m_stepsize* 108

5.11 The average of the objective function minimised using the PSO algorithm and the proposed metric-based approach with different levels of investment ($m \in [0.5, 1, 2, 4, 8]$). The objective function is the expected initial service disruption in percentage of the normal service level for a given weather profile. m indicates the investment level. $m = 2$ means the amount of investment equals to that required to increase the design load for all assets in the network by 2°C. 109

5.12 Histogram showing the calculated Kendall's τ-b coefficient between the solutions found by the PSO algorithm and the iterative metric-based method. 110

5.13 Combine the solutions by averaging along each edge. 110

5.14 Bar plots showing the combined resource allocation solutions in relation to their latitude. Each bar represents an edge in the network, and the length of the horizontal bar represents the amount of increase in its design load. The bar plot organises the edges by the latitude of their centroid. The south-most edge is at the bottom of the plot. The y-axis is asset's rank by latitude and is not actually its latitude. 112

5.15 Correlation between the solution and the latitude of the edges. 113

5.16 The decrease of probability of failure against increasing $\Delta\mu_i$ for when the local tasmax is 38°C (red) and 28°C (blue). 113

5.17 Correlation between the solution and the unit cost of the edges. 114

5.18 Correlation between the solution and the flow throughputs of the edges. 115

5.19 Maps showing the resource allocation solutions obtained from optimising toward the 5 hottest day each year for the years 2051-2100. Width of the lines indicates the increment in design load $\Delta\mu_i$ for the asset. m is the variable used to quantify the amount of investment. $m = 2$ means the amount of investment equals the amount required to increase the design load for all assets in the network by 2°C. 116

5.20 Maps showing the resource allocation solution obtained from optimising toward the 250 representative examples days selected from the summer days of the year 2051-2100. Width of the lines indicates the increment in design load $\Delta\mu_i$ for the asset. m is the variable used to quantify the amount of investment. $m = 2$ means the amount of investment equals the amount required to increase the design load for all assets in the network by 2°C. 117

Acronyms

2D Two-Dimensional.

3D Three-Dimensional.

AADPT Annual Average Daily Passenger Trips.

AR5 The Fifth Assessment Report.

AR6 The Sixth Assessment Report.

CDF Cumulative Distribution Function.

CORDEX Coordinated Regional Climate Downscaling EXperiment.

CRT Critical Rail Temperature.

ESGF The Earth System Grid Federation.

EURO-CORDEX European Coordinated Regional Climate Downscaling EXperiment.

GB Great Britain.

GDP Gross Domestic Product.

HPC High Performance Computing.

IPCC Intergovernmental Panel on Climate Change.

LOS Loss Of Services.

PSO Particle Swarm Optimisation.

RCP8.5 Representative Concentration Pathway 8.5.

SLSQP Sequential Least SQuares Programming.

SQP Sequential Quadratic Programming.

UK the United Kingdom.

Chapter 1

Introduction

1.1 Background

Climate change is a challenge for all infrastructure sectors (Pörtner et al., 2022; Masson-Delmotte et al., 2021; Betts et al., 2021). The global mean surface temperature has increased by 1.09°C from 1850-1900 to 2011-2020 based on the conclusion in The Sixth Assessment Report (AR6) (Masson-Delmotte et al., 2021). Figure 1.1 shows the historical and projected change of four key indicators of global climate. The global warming is likely to surpass the threshold of 1.5°C above pre-industrial level in the early 2030s or earlier under a fossil fuel-rich development (Masson-Delmotte et al., 2021).

It is well evidenced by climate science research that a changing climate causes changes in the frequency and intensity of some weather extremes. The frequency and intensity of hot extremes have increased and will continue to increase. As global warming increases, global land precipitation will increase with substantial regional variations, seasonal contrasts and variability. Heavy precipitation, extreme storms, and high-category tropical cyclones are projected to become more frequent in general with increased intensity at the global scale (Seneviratne et al., 2021). Climate change causes significant changes in the spatial and temporal distributions of most of the climate variables, which could lead to more frequent and severe extreme events happening at new times and new locations.

The changes in weather extremes pose challenges to infrastructure systems of all kinds. Infrastructure systems are collections of physical structures such as roads, bridges, railways, water pipes, sewer lines, transmission lines, transmission towers, and telecommunication cables. Those physical assets are designed, built and operated in compliance with design codes and regulations set on historical meteorological data and are expected to cope with certain magnitudes of external weather parameters. However, the parameters they were designed for have been largely surpassed by the scale of recent extreme weather events caused by climate change, such as Storm Arwen in December 2021(BBC, 2021) and the July heatwave in 2022 (Met Office, 2022), and are likely to be further exceeded in the future as the current climate change trajectory is only

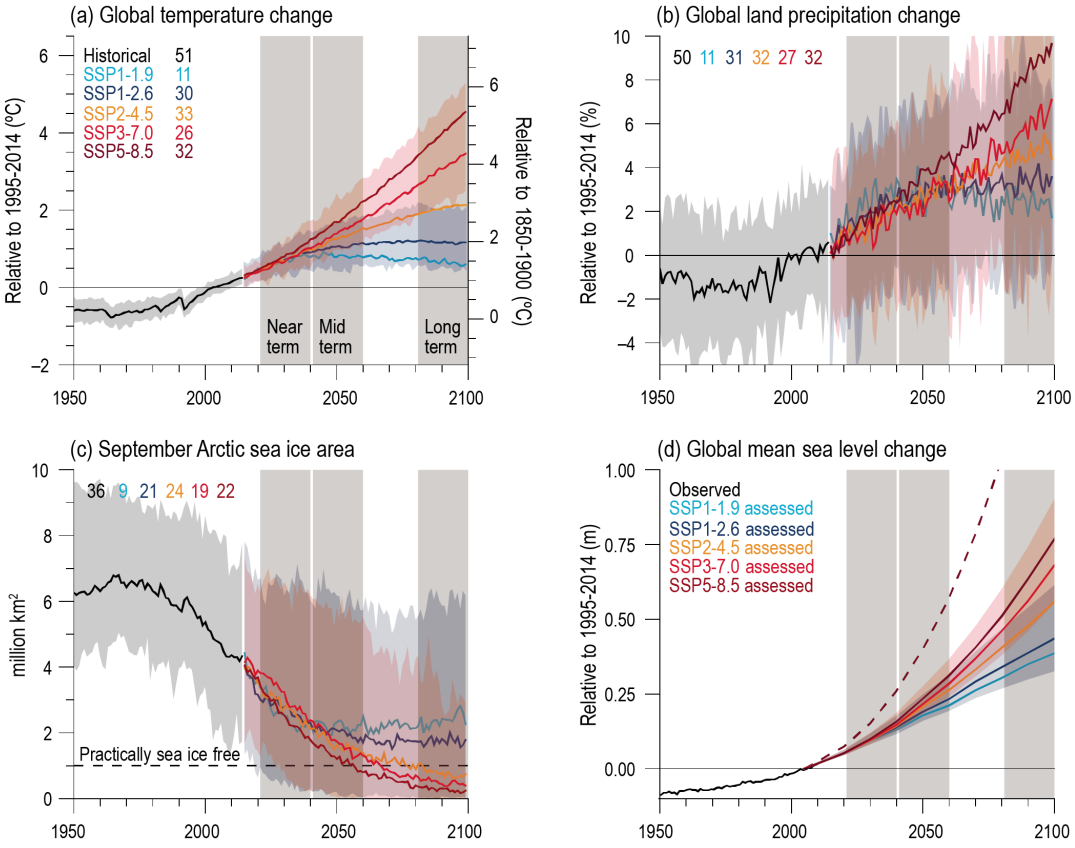


Figure 1.1: Four indicators of global climate change from Masson-Delmotte et al. (2021)

likely to result in an increased frequency and intensity of weather extremes.

There is no shortage of evidence demonstrating the impacts of extreme events in the past few years (Met Office; National Weather Service; Douris and Kim, 2021). There has been an increasing body of literature on estimating the impacts of climate change on infrastructure systems. The most straightforward and plainest way for climate change impact assessment is to count the number of days that a certain threshold (e.g. critical air temperature for railway tacks, and maximum wind speed for transmission towers) of a given climate variable (e.g. daily maximum near-surface air temperature) for a hazard mechanism of interest (e.g. rail buckling in rail transport systems) will be exceeded in the future. A progress is to project aggregated or further processed maps of climate variables on maps of infrastructure assets and estimate the number of, or the total length of, physical infrastructure assets under risk. However, beyond that, there is a lack of efforts on quantifying climate change impact in the form of service disruptions, (e.g. number of train journey cancelled and number of people left without electricity). It drives the need to translate climate change projects into projections of system-level service disruption, moving beyond estimations of component level asset failure (Figure 1.2).

Figure 1.2: A simplified process on how the climate system impacts the infrastructure systems

Infrastructure systems are crucial for productivity, economic growth, and quality of life (HM Treasury, 2020). Recent events have highlighted the precarious state of the national infrastructure systems. Given the understanding that current infrastructure systems may not have the ability to withstand the future climate, adaptation plans are much needed to reduce climate change impacts on infrastructure systems. Sharing the same general understanding, the UK government committed £640 billion of gross capital investment in the national infrastructure before 2024-25 (Betts et al., 2021) and 1.1% - 1.3% of the GDP each year from 2025 to 2055 (Treasury, 2021). Infrastructure system decisions made today underpin their behaviour in the next 50 years. With the changing climate, the allocation of climate change adaptation resources and infrastructure capital investment should factor in future climate possibilities and optimise toward plausible future patterns of extremes.

1.2 Research Aim and Objectives

Overall, this research aims to first understand and then mitigate the impact of climate change on large infrastructure systems. Motivated by the discussions above, two research questions are asked:

1. How can the magnitude of future infrastructure service disruption due to climate change be estimated, accounting for corresponding uncertainties?

Upon answering the above question and quantifying the magnitude of service disruption, the next question is:

2. How should resources be allocated in the most efficient manner to upgrade infrastructure assets so that the forthcoming service disruptions can be minimised?

There exist assessments carried out to quantify the level of service disruption under simulated extreme events (e.g. assessing the impact of a 1 in 100-year flooding event on an urban transportation network). These assessment frameworks often first simulate failures in the system, then simulate how the system responses and finally quantifies the consequences (e.g. loss of service). In these existing assessment framework, the concept of resilience, along with robustness and vulnerability, is employed to measure the system's ability to provide services under climatic hazards. Such resilience assessment frameworks can be used to estimate the magnitude of service disruption for a given hazard. As discussed above, there exist climate change impact assessments that estimate the scale of asset failure, which has the potential to be used as input to resilience assessment framework to estimate the scale of service disruption. An approach to simulate failure scenarios through future weather profiles generated by climate models, which can be integrated into a resilience assessment, is hence needed.

Resource allocation problems for networked systems are of interest to two separate research fields. The analysis of network structure and graph topology leads to the development of metrics that

can be used to measure the relative importance of the network's components. Such metrics can direct resources to the most critical places in the network. However, existing metrics appear to focus more on the network's topological structure and overlook its supply-demand feature and vulnerability variations associated with climate hazards. At the intersection of mathematical optimisation and operational research, the resource allocation problem can be formulated as an optimisation model with the objective function linked to the infrastructure system's resilience. They are deceptively easy to formulate. However, given the large number of assets in any real-world infrastructure system, the optimisation problems often turn out to be high-dimensional and soon become computationally intractable. A tractable strategy to allocate resources effectively and efficiently for large-scale infrastructure systems is needed.

In this thesis, the aim will be achieved and the research questions will be addressed by meeting the following objectives:

1. To review literatures in relevant fields (Chapter 2);
2. To propose a method to estimate the magnitude of future infrastructure service disruption due to climate change accounting for uncertainties (Chapter 3);
3. To develop an method to allocate resources in the most efficient manner to upgrade infrastructure assets so that the forthcoming service disruptions can be minimised (Chapter 4);
4. To benchmark the proposed methods with case studies (Chapter 3 and Chapter 4).

1.3 Research Contribution

By addressing the research questions and carrying out case studies, this thesis offers the following contributions to knowledge:

1. random failure models tend to overestimate the network's resilience and the railway passenger transportation network response differently to random, targeted and weather-related failure scenarios;
2. the scale of the system-level service disruption degrades non-linearly with the magnitude of the disruption and the magnitude of disruption scale non-linearly with increasing temperature;
3. a low level of investment leads to higher inequality in asset upgrades. It implies that when resources are limited, concentrating them on fewer but more critical assets could be more beneficial than spreading the resources out;
4. alongside other factors such as the topological importance and functional importance, the climate-related failure probability plays an indispensable role in directing the allocation of resources in the most efficient possible way.

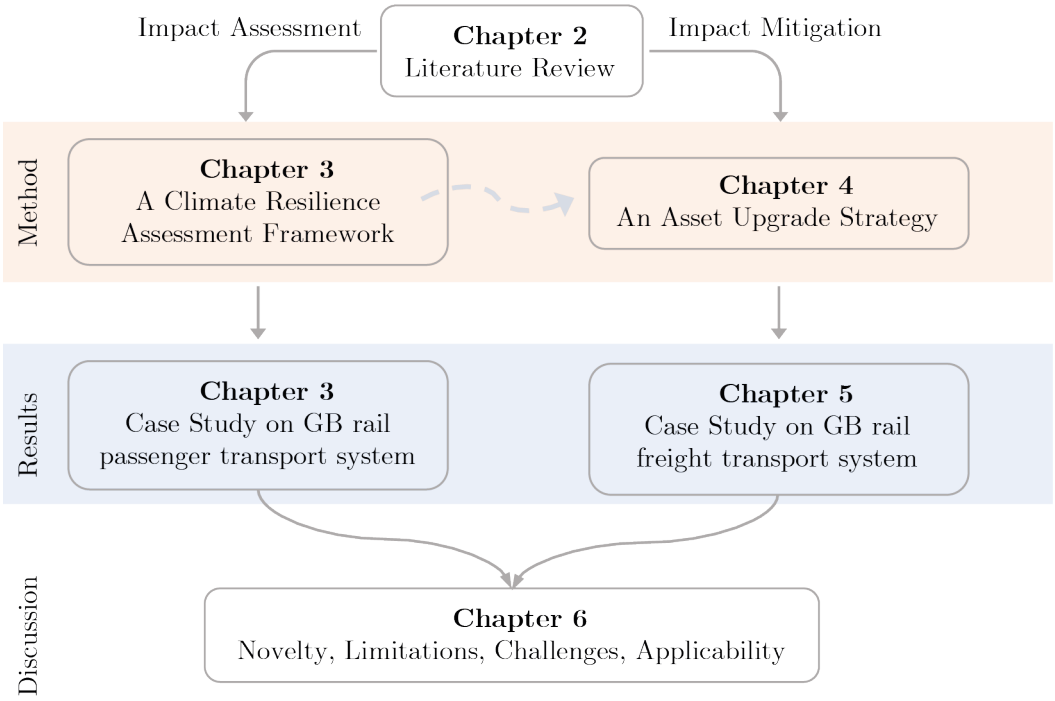


Figure 1.3: Chapter Structure of this Thesis

1.4 Thesis Structure

Figure 1.3 shows the high level structure of the thesis. The method proposed to address the first research question is presented in Chapter 3, which also includes a case study demonstrating the proposed climate resilience assessment framework. The method proposed to answer the second research question is presented in Chapter 4 with the corresponding case studies on resource allocation problems presented in Chapter 5. The thesis is organised as follows:

Chapter 2 - Literature review

> This chapter first discusses the impact of climate change on infrastructure systems and the quantification of such impacts, pointing out the necessity of a systemic climate resilience assessment framework that uses representative future failure scenarios to address system-level service disruption. This chapter then reviews two research approaches relevant to the resource allocation problem - node/edge ranking in network science and mathematical optimisation, pointing to a need to bring the two methods together. The research gaps and objectives are identified at the end of this chapter.

Chapter 3 - A systemic climate resilience assessment framework

> This chapter presents a novel probabilistic resilience assessment framework that generates failure scenarios and network disruptions using weather profile data from climate prediction models with component-level fragility functions. A case study is then carried out to quantify the resilience of Great Britain's railway passenger transport system to high-temperature-related track buckling under the Representative Concentration Pathway 8.5 (RCP8.5) climate change scenario.

Chapter 4 - Asset upgrade strategy towards climate resilience

This chapter considers the problem of finding the optimal or best possible resource allocation between infrastructure assets to minimise the system's service disruptions under plausible future extreme weather events with a finite budget. It first presents the formulation of the resource allocation problem as a constrained optimisation problem. The formulated problem is then approached with a metric-based method and an analytical method, which will be shown to converge to the same expression and lead to the proposal of a novel network metric that can efficiently direct resource allocation.

Chapter 5 - Network metric-based resource allocation

This chapter presents three case studies that implement and further develop a ranking-based resource allocation strategy using the network metric proposed in the previous chapter. The proposed method is applied to Great Britain's core freight network, showcasing its ability to handle resource allocation problems for large-scale infrastructure systems effectively. It outperforms a meta-heuristic optimisation algorithm in both time efficiency and solution optimum.

Chapter 6 - Discussion

This chapter presents a discussion on the novelty, limitations, implementation challenges and general applicability of the developed climate resilience assessment framework and the proposed resource allocation strategy using a novel network metric.

Chapter 7 - Conclusion

Concluding remarks and recommendations for future works are presented in this last chapter.

Chapter 2

Climate Change and Infrastructure Systems

2.1 Chapter Introduction

This chapter contains the examination of three bodies of literature. Section 2.2 briefly discusses how climate change impacts infrastructure systems in general and how this impact has been quantified in the literature. Section 2.3 introduces the concept of resilience and reviews existing resilience assessments in terms of their system modelling, disruption generation method, failure determination approaches and asset recovery settings. Section 2.4 reviews works on using network metrics as network component importance measurements to aid resource allocation and optimisation models formulated for resource allocation in infrastructure systems. Finally, Section 2.5 discusses the existing methods, identifies two research gaps, and sets the research objectives to address the research questions. [1]

2.2 Climate Change and Infrastructure Systems

2.2.1 Climate Hazards to Infrastructure Systems

According to IPCC, 'hazard' is defined as *'the potential occurrence of a natural or human-induced physical event or trend that may cause ... damage to infrastructure ...'* (Pörtner et al., 2022). Based on this definition, climate-related hazards to infrastructure systems are the potential occurrence of weather events where the weather parameters may surpass the designed capacities of physical infrastructure assets which may then trigger asset failures and pose negative impacts on infrastructure systems. Climate-related hazards have the potential to cause disruption via various mechanisms and to different extents. Take road and rail transport for example, the most

[1] The current chapter contains materials that were previously prepared for the following: Li, Q., Punzo, G., Robson, C., Arbabi, H. and Mayfield, M., 2022. A Systematic Approach to Climate Resilience Assessment of Infrastructure Networks.

harmful climate-related hazards are snowfall, cold spell, heatwave, heavy rain, and wind gusts (Vajda et al., 2014).

Snowfall: Light snowfall with low surface temperature may increase road slipperiness and increase the number of road accidents. While heavier snowfall may significantly reduce road surface friction and visibility, in which case, road closure is very likely. For rail transport, heavy snowfall mostly impacts the system via snow and ice built-up on the tracks and operation degradation caused by failure of overhead electrical lines (Ludvigsen and Klæboe, 2014).

Cold spell: Low temperature increases the road slipperiness when there is snow or rain. Temperature below 7°C also reduces the effectiveness of road deicing salt, leading to higher accident rate and potential road closure (Vajda et al., 2014). Rail line shutdown may happen due to reduced grip of wagon brakes on slippery track surface (Ludvigsen and Klæboe, 2014). Extreme low temperature may also cause damage to track, switch and railcar, leading to longer term effects (Greenham et al., 2020).

Heat wave: Impacts of extreme high temperature mostly relate to structural damages: buckling of road surface and railway tracks, and thermal expansion of bridge joints (Palin et al., 2021). Other common effects of high temperature include softening of road pavement and rail equipment failure (Vajda et al., 2014). Though high temperature is considered to be of relatively minor influence, with climate change, prolonged periods of summer heat waves that return every few years may still cause profound effects to the infrastructure assets in longer terms (Philip et al., 2021).

Heavy rain: The direct impact of heavy rain is reduced visibility and increased slipperiness. Heavy precipitation triggered flash flood further impacts the system more significantly: inundation/flood water overtopping of road surfaces, rail track and bridges, scour of bridges and railway track embankment, and damage to roadbed (Tsubaki et al., 2016; Lam et al., 2018).

Wind gust: Fallen trees brought down by strong winds can block roads and railway tracks. Fallen tress and other objects such as plastic bags can also damage overhead lines of the railway network, causing power and signalling failures (Ludvigsen and Klæboe, 2014; Jaroszweski et al., 2015).

Climate variables, such as temperature, wind speed, precipitation and air pressure, exceeding certain thresholds causes those climate-related hazards. Figure 2.1 shows a conceptual map of the complex interactions between climate variables, climate-related hazards, and their impacts on the road infrastructure and transport. A climate hazard could be the result of one or more climate variables. A climate hazard could impact one or more aspects of a given infrastructure systems. For example, high temperature disrupts the operations of the railway from multiple aspects such as the increased probability of track buckling and overhead line sagging. Further, multiple climate hazards acting together could lead to compound events. For example, drought when combined with extreme heat and low humidity increases the probability of wildfire outbreaks. The interaction between the climate system and infrastructure systems are inherently complex. In principle, it is the extreme values of the climate variables that causes the climate-

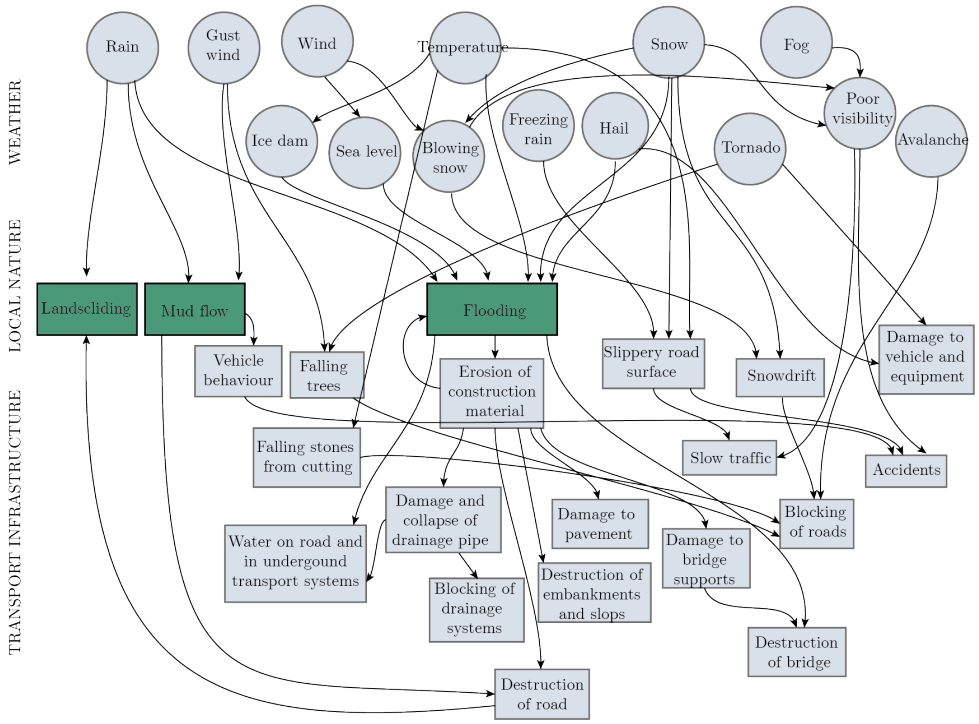


Figure 2.1: Conceptual map of interacting weather phenomena impact on road infrastructure and transport. Recreated from Leviäkangas et al. (2011).

related hazards, which may then cause failures and disruptions in the infrastructure systems.

2.2.2 Impact of Climate Change on Infrastructure Systems

As discussed in the previous section, climate-related infrastructure asset failures and service disruptions are caused by weather extremes exceeding infrastructure assets' design loads. In order to understand the impact of climate change on infrastructures systems, we first need to understand the relationship between climate change and the weather. The difference between weather and climate is the timespan. The Intergovernmental Panel on Climate Change (IPCC) (IPCC, 2013) defines climate *'as the average weather, or more rigorously, as the statistical description in terms of the mean and variability of relevant quantities over a period of time'.* The term 'weather' describes the atmospheric condition over a short period of time, from hours to days, in terms of temperature, humidity, precipitation, wind and etc. (IPCC, 2013). In short, climate describes the overall weather pattern over a longer period of time and a changing climate means a changing weather pattern. The impacts of climate change on infrastructure systems are posed through the change of occurrence patterns of weather extremes.

It is well established that climate change will lead to an increasing frequency and intensity of weather extremes (IPCC, 2013). Houghton et al. (2001) uses a schematic illustration to show the effects of increased mean, variance, and combined mean and variance of temperature distribution on temperature extremes (Figure 2.2). The empirical distribution of temperature approximately fits a normal distribution. Therefore a change in the mean temperature leads to new highs of

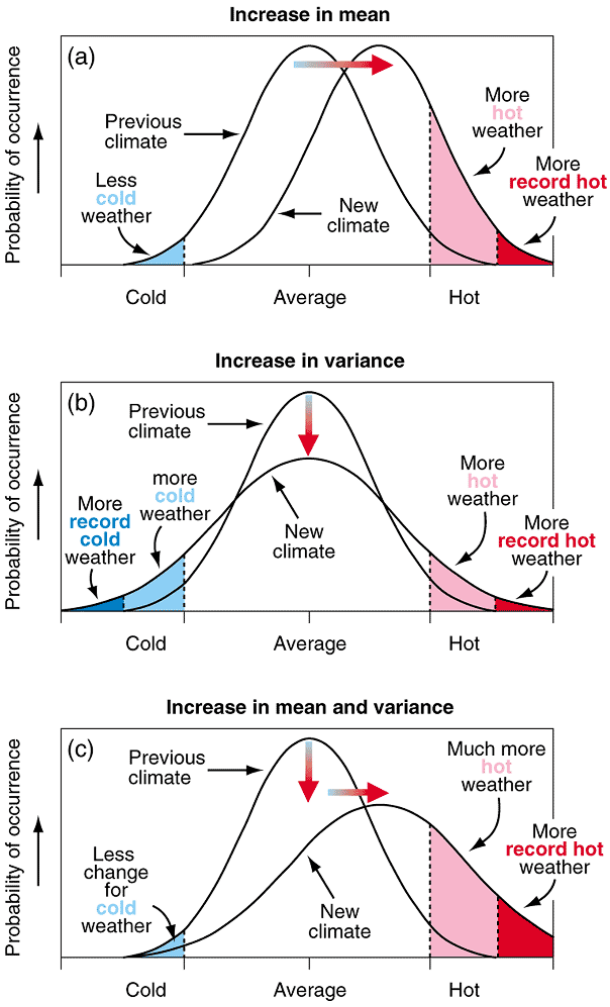


Figure 2.2: Schematic showing the effect on extreme temperatures (a) the mean temperature increases, (b) the variance increases, and (c) when both the mean and variance increase for a normal distribution of temperature (Houghton et al., 2001)

hot extremes and increased probability of passing a specific temperature threshold (for example, temperature for safe operation of some infrastructure system assets). An increase in variance of the distribution with the mean remaining the same, leads to an increase in frequency of extreme hot and cold as well as record-breaking high and low temperature. A combination of both, depending on the magnitudes of change of each parameter, could result in more hot extremes and less cold extremes with an increase in heatwaves.

A key message from Figure 2.2, which is also evidenced by other researchers (Rummukainen, 2012; Zwiers et al., 2013; Overland et al., 2016; Ohba and Sugimoto, 2020), is that the change in frequency and intensity of certain types of weather extremes could be disproportional to changes in the climate system. The most recent IPCC assessment report, AR6, point out that even relatively small increases in global mean surface temperature (+ 0.5°C) can lead to statistically significant changes in the occurrence of extreme weather events with larger changes for more extreme and rare events.

The change in the likelihood of surpassing a certain hot extreme does not scale linearly with

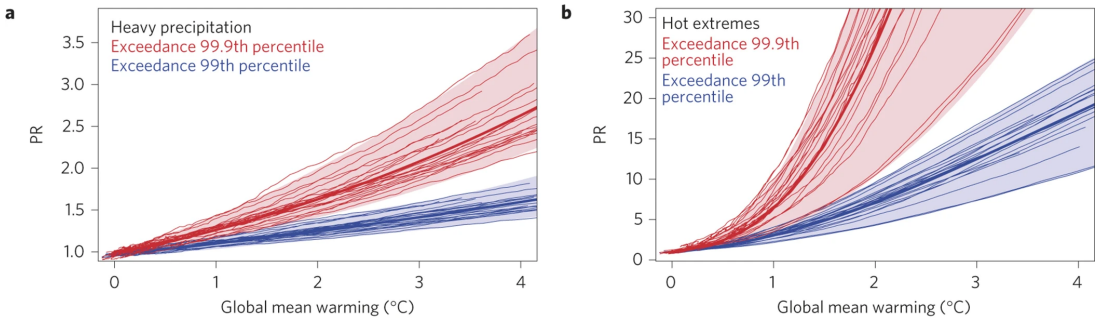


Figure 2.3: Probability ratio (PR) of exceeding the 99th (blue) and 99.9th (red) percentile of pre-industrial daily (**a**) temperature and (**b**) precipitation at a given warming level, averaged across land. The probability ratio is defined as the probability of exceeding a certain quantile with a given level of global warming over the probability of exceeding it during the pre-industrial control period (Fischer and Knutti, 2015). Figures from Fischer and Knutti (2015)

the extent of global warming, as reported in AR6 (Seneviratne et al., 2021). The frequency of a present-day 20-year return period heat extreme is projected to increase by 80% with 1.5°C global warming and by 180% with 2.0°C warming. The frequency of a present-day 100-year return period heat extreme is predicted to increase by 200% with 1.5°C global warming and by more than 700% with 2.0°C warming. In summary, rarer events are likely to see larger changes in its frequency.

AR6 (Seneviratne et al., 2021) concludes that the global averaged annual maximum of one-day precipitation is projected to change linearly with the degree of global warming. The intensity of a 50-year return period event increase at a rate close to 7% per 1°C of global warming. The frequency, however, is project to increase non-linearly with larger changes for more rare events. At a global warming level of 1.5°C, there is a projected 10% increase in the frequency of a 20-year return period extreme precipitation and 20% increase for a 100-year event. At a global warming level of 2.0°C, the projected increase is 22% and more than 45% respectively for 20-year and 100-year events. At a global warming of 4°C, the increase in frequency for 10-year and 50-year events increases by 100% and 200%.

Beyond temperature and precipitation, projection on other climatic phenomena, such as flooding, droughts and storms as less quantifiable and of various levels of confidence due to the inherent complexity of their processes (Seneviratne et al., 2021). With global warming, the percentage of global land area affected by flooding is predicted to increase (medium confidence), primarily due to changes in heavy precipitation. Though it appears to be controversial, regions affected by droughts will also increase with global warming. Further, the total number of tropical cyclones is projected to decrease as a result of a significant reduction in the number of weak tropical cyclones (weaker than or equal to Category 1). However, the frequency of intense tropical cyclones (Category 4 and 5) are projected to increase with an increase in intensity measured by maximum wind speed and central pressure. In summary, the frequency, intensity and variability of extreme weather events are projected to increase.

AR6 also points out five main challenges from climate change. First, the future will witness the

occurrences of extreme weather events with magnitudes that surpass historical records. Current infrastructure systems, which comprise physical assets that are designed, built and operated in compliance with design codes and regulations set on historical meteorological data, may not have the ability to withstand the future climate. Second, extreme weather events, will also occur at increased frequency, meaning the design thresholds for current infrastructure systems will be surpassed more frequent than expected by their design and operational standard. Frequently surpassing load-bearing metal structures' design load could lead to fatigue failure and shorter life expectancy (Hulse and Cain, 2018).

Third, regions may experience types of extremes that they have never encountered before and infrastructure systems in those regions may face types of extreme that they were not originally designed to handle. Road and rail infrastructure in the arctic region is facing huge challenges due to permafrost thawing and ground collapse (Vincent, 2020). Droughts and floods occurring in new regions disturb the electricity generations and threaten local energy security, in particular for regions that rely heavily on hydropower (Shadman et al., 2016).

Fourth, extreme events may also occur at new times of the year. Climate-related asset failure may occur earlier than infrastructure management company anticipated and resources are not readily available. For example, in the railway industry, heat-related incidents are more likely to happen in the early summer as temperature start to rise. With global warming, the rising of temperature may start earlier in the year before scheduled seasonal maintenance activities (Ferranti et al., 2016). Finally, compound events may see new combinations of type, timing, and/or severity, aggravating the impact and consequences of extreme events. Current infrastructure systems may not have the ability to withstand the future climate, characterized by more frequent and intense weather extremes with new patterns. What is resilient to the present-day climate may be vulnerable to the future climate. A comprehensive understanding of how future weather hazards impact infrastructure systems is required.

2.2.3 Climate Change Impact Assessment

Since the publication of the Intergovernmental Panel on Climate Change (IPPC) Fifth Assessment Report (IPCC, 2013) and the open access to climate model output data, there has been an increasing body of literature on assessing the impacts of climate change on infrastructure systems. For example, the third Climate Change Risk Assessment (Betts et al., 2021) assesses the future flood risk to Great Britain and estimates the number of assets exposed by overlapping the flood risk maps with the geographical maps of the assets and identifying those likely to surpass specific indices or risk thresholds. Other impact assessments appearing in the literature have been carried out in a similar manner. They rely on estimating the likelihood of surpassing certain design or operational thresholds under several climate change scenarios (Bubeck et al., 2019; Koks et al., 2019; Sanchis et al., 2020).

For the classification and determination of the impact of climate hazards, a two-tier approach is proposed in EU-CIRCLE, where the approach first distinguishes direct and indirect impact. Di-

rect impacts refer to consequences related to the infrastructure system itself, including complete or partial damage to physical infrastructure assets, deviation of performance from the fully functional level, and connectivity loss. Indirect impacts refer to those received by the society that is served by the infrastructure systems, such as causalities, community isolation (both physical and in terms of communications), and economic losses. By this classification, the majority of existing climate change impact assessments, including Bubeck et al. (2019), Koks et al. (2019), Sanchis et al. (2020) and Betts et al. (2021), are on the physical damage level.

Under the scope of climate change impact assessments, functional damages are rarely analysed, often just viewed as a consequence. Works related to functional damage and system-level loss of service, in particular, are comparatively more limited in number (Mostafavi, 2018; Wilkinson et al., 2022). Limiting impact assessments to the estimation of physical damages, or even component-level functional loss overlooks the complex interdependencies of infrastructure systems (Rinaldi et al., 2001). Infrastructure systems are well acknowledged as complex coupled systems, the behaviour or response of which is distinct from the combined behaviour or response of its components (Punzo et al., 2018). Impact assessments on physical damage level certainly provide valuable insights into the magnitude of climate-change-related disruptions. However, the complex dynamics characterizing infrastructure systems' responses are such that system-level effects cannot be derived in a straightforward manner from the component level. Therefore, a systematic understanding of how more extreme weather events may cause a hazard and impact infrastructure system's ability to maintain services is required.

2.2.4 Brief Summary

During extreme weather events, assets can fail when the local weather parameter exceeds the threshold the assets are designed to withstand. Climate change causes changes in the temporal occurrence and spatial patterns of such extreme weather events alongside increased intensity. Utilising the extensive amount of climate projection data, data analyses were carried out to estimate the increase in the frequency of a given extreme weather event (e.g. heavy precipitation) with global warming. There also exist works that estimate the frequency of a specific design threshold for a given type of infrastructure assets being surpassed and the number of infrastructure assets under a given level of risk. However, there is a lack of work that estimates the magnitude of system-level service disruption. A framework is needed to progress climate change impact assessment on infrastructure systems from component-level asset damage estimation to system-level functional loss estimation.

2.3 Resilience as A Tool for Climate Change Impact Assessment

2.3.1 The Concept of Resilience

Existing works on system-level responses to disruptive events are often cast within research in resilience, as well as the related areas of vulnerability and robustness. In general, resilience is a concept associated with the way a system responds to adverse events, system disruption, disturbance, or shock. The concept of resilience was first introduced to ecological systems by Holling (1973). Since then, it has been widely applied to many disciplines and interpreted through different definitions (Table 2.1).

Table 2.1: Example Definitions of Resilience

Definition	**System**	**Reference**
'A measure of the persistence of systems and of their ability to absorb change and disturbance and still maintain the same relationships between populations or state variables.'	Ecological system	(Holling, 1973)
'The ability of transportation systems to experience a potentially damaging event and return to a healthy state of operations in a reasonable period of time after that event.'	Metro	(Chan and Schofer, 2016)
'Network resilience is defined as the expected fraction of demand that can be satisfied post-disaster.'	Freight transport	(Miller-Hooks et al., 2012)
'The adaptive capability of the supply chain to prepare for unexpected events, respond to disruptions, and recover from them by maintaining continuity of operations at desired levels of connectedness and control over structure and function'	Supply chain	(Spiegler et al., 2012)
'The ability for the system to absorb the consequences of disruptions to reduce the impacts of disruptions and maintain freight mobility.'	Freight transport	(Ta et al., 2009)
'Resilience is the ability of a system to prepare for, mitigate, adapt to, and/or rapidly recover from the effects of endogenous and exogenous "disruptive" events.'	Complex Engineering System	(Punzo et al., 2018)

Some of the definitions focus on breaking the concept of resilience down into multiple system capacities. For example, Punzo et al. (2018) define resilience as a system's ability to 'prepare for, mitigate, adapt to, and/or rapidly recover from' disruptive events. Resilience defined by EU-CIRCLE emphasizes the system's anticipative, absorptive, coping, restorative, and adaptive capacity. The UK National Infrastructure Commission lists 'anticipate, resist, absorb, adapt and transform' as key aspects of infrastructure resilience (National Infrastructure Commission, 2020). While 'robustness, resourcefulness, rapid recovery, adaptability' are identified as the four main features of resilience by the National Infrastructure Advisory Council of the USA (Berkeley

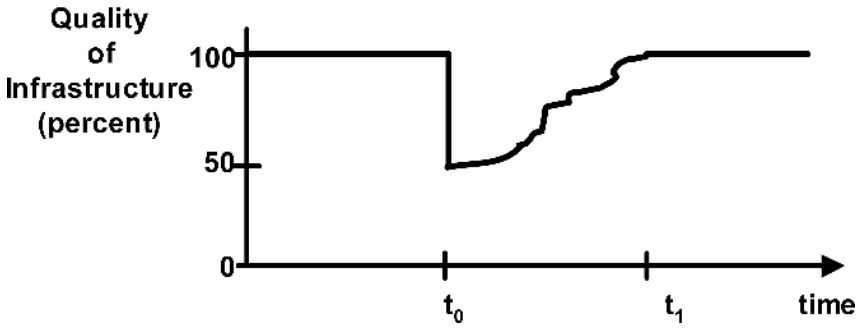


Figure 2.4: The conceptual definition of the measure of resilience proposed by Bruneau et al. (2003). The curve shows the variation of the quality of the infrastructure (Q(t)) from the time of the event happening ($t=t_0$) to the time of full recovery ($t=t_1$). The quality of infrastructure ranges from 0% (no services) to 100% (fully functional) and is measured as the proportion if infrastructure service delivered. Figure from Bruneau et al. (2003).

et al., 2010).

Though there is no consensus on the definition of resilience, the resilience of a system to a given disruption is often quantified as the cumulative service loss between the time of the disruption event happening and the time of full recovery. Bruneau et al. (2003) creates the resilience triangle in order to measure a community's resilience to earthquake (Figure 2.4). The resilience triangle is formed by a conceptualised resilience curve and the line of system's normal performance level. Area of the triangle represents the total loss of infrastructure functionality and the amount of service disrupted. Shape of the resilience curve reflects the system's course of failure and recovery, as a joint result of those resilience related capabilities including robustness, redundancy, resources, and rapidity. Under such a quantification framework, resilience is defined as the cumulative service loss from the time of disruption to the time of full recovery.

The concept of robustness originates in network science, where robustness is associated with the network's ability to stay connected when a fraction of nodes/edges is removed from the network and measured as extent of reduction in network connectedness with successive nodes/edges removal (Albert et al., 2000). When applied to infrastructure systems, instead of network connectedness, robustness sometimes is linked to the system's ability to maintain the continuity of its service. For instance, Faturechi and Miller-Hooks (2015) defines the transportation system robustness as "the ability of a system to continue in operation and thus maintain some level of functionality even when exposed to disruption". By their definition, the concept of robustness relates to the system's ability to resist service drop but does not concern the system's ability to return to its pre-disruption performance level.

From reviewed literature, vulnerability can be interpreted in two ways. The first interpretation of vulnerability is a system-level property that is associated with both the probability of the occurrence of disruptive events and the extent of resulted negative impacts (network damage or service loss) (Johansson and Hassel, 2010). Berdica (2002) defines transport system vulnerability as 'a susceptibility to incidents that can result in considerable reductions in road network serviceability'. Vulnerability in this context relates to the definition of risk - the product of

probability of events happening and magnitude of negative consequences. Both robustness and vulnerability consider magnitude of negative consequences but robustness only quantifies the adverse effect (service loss or network damage) for a given disruptive event without considering the probability for such event happening.

The second interpretation of the vulnerability of a system refers to the critical components of a system. A component is said to be a vulnerability of a system when its removal or failure can cause significant consequences, which may even cause the whole system to fail. Application of vulnerability in this context to transportation systems can also be spotted. For example, Bababeik et al. (2017) develop a vulnerability analysis framework to identify critical links in a railway network. Pant et al. (2016) study the vulnerability of an interconnected railway, power, water and signalling network to identify critical supporting systems to passenger travelling network.

Overall, resilience is a more holistic view of the system's ability to withstand external disruptive events than robustness or vulnerability. Utilising the concept of resilience with its multiple dimensions (such as anticipation, robustness, rapidity, and adaptation), it is possible to set up a systematic framework for assessing, analysing, and enhancing the climate resilience of infrastructure systems.

2.3.2 State of the Art: Resilience Assessment Frameworks

Main steps in a quantitative resilience assessment using simulation approach are system modelling, disruption generation, failure determination, and asset recovery. The rest of this section follows this structure and reviews approaches and techniques used. Given that the large number of types of infrastructure and the vast body of literatures on infrastructure resilience assessments, the scope of literature reviews is narrowed down to transportation network when reviewing types of system models. When reviewing methods used in disruption and failure generations, methods reviewed expands electricity networks because such methods are not available in the transportation network literature.

System models

A resilience assessment often starts with modelling real-world infrastructure systems as networks. Network models in literature can be categorized into topological models, flow models and combined models. In a topological model, physical assets of infrastructure systems, such as rail stations, transmission lines, and airports, are modelled as nodes or edges in the network (Gastner and Newman, 2006; Von Ferber et al., 2009; Lin and Ban, 2013; Robson et al., 2021). The topological models are often used to study the statistical properties and spatial structure of the infrastructure networks. For example, in Von Ferber et al. (2009), the public transportation networks of 14 major cities worldwide were modelled as graphs where each nodes represents a station and an edge indicates there exists public transports linking the two stations. They found

the networks studied showing a strong small-world structure with high clustering coefficients and small average length of shortest paths. These models emphasise on the topological structure of the system but lack the ability to capture the functional aspect of the infrastructure system: where the demands are, where the supplies are and how the demands can be met.

A flow model has its emphasis on the network structure of the flow of service or products between infrastructure sectors or key entities. For example, researchers have modelled the supply chains as flow models to analysis the product flows, their equilibrium and price dynamics. Nagurney et al. (2002) developed an equilibrium model of competitive supply chain network model (Figure 2.5). The manufactures, retailers and consumers were modelled as three groups of nodes, linked by edges representing transportation of products. Using the proposed network model, they demonstrated the computation of the equilibrium price and product flow pattern. Hearnshaw and Wilson (2013) revealed that the properties of an efficient supply chain network can be mirrored by a scale-free network. Supply chain resilience studies in general concern the system's ability to sourcing supplies to satisfy the demands despite disturbances (Hosseini et al., 2019). However, those models fail to consider that the transportation of those commodities relies on the functioning of physical infrastructure assets, such as ports, highways and trains. The interactions between the physical infrastructure assets and the flow of services are missing.

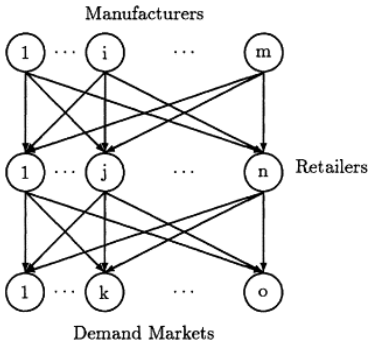


Figure 2.5: The network model of the supply chain in Nagurney et al. (2002), which consists of m manufactures (the top row of nodes), n retailers (the middle row), and o demand markets. Figure from Nagurney et al. (2002).

Recent works on system-of-systems, interdependent, or interconnected networks adopt a combined topological and flow model (Pant et al., 2016; Goldbeck et al., 2019). Such models have separated asset and flow layers representing the physical infrastructure assets and the service provided. Inter-layer dependencies are used to link the two layers together. Figure 2.6 shows a simple example used in Goldbeck et al. (2019) to illustrate the system representation an dependency types. The model is designed on the basis that network flows may require the functioning of multiple assets of an infrastructure sector, which could then further rely on more assets that belongs to different infrastructure sectors. The separation makes it possible to model various types of interdependencies independently.

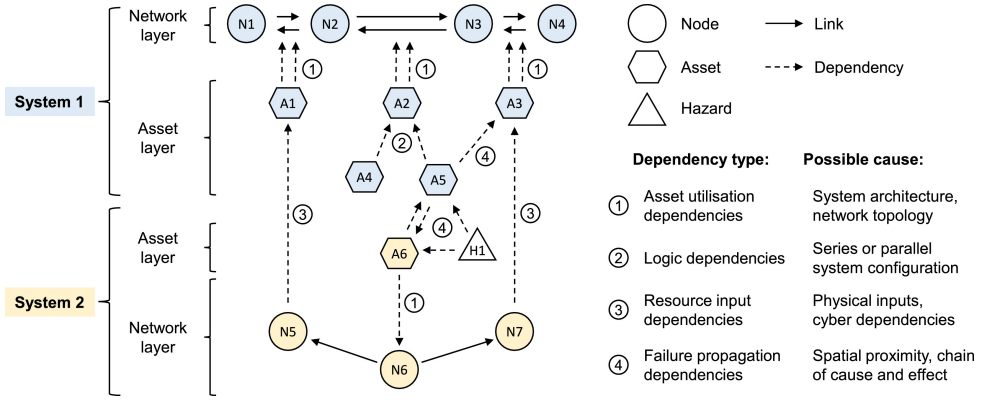


Figure 2.6: A simple example of two interdependent systems with four types of dependency. For example, the flow between N1 and N2 utilises asset A1 in System 1. Figure from Goldbeck et al. (2019).

Disruption generation

In the field of network science, disruptions are mostly simulated as strategic removals of network components (Berche et al., 2009; Pagani et al., 2019). This type of network component removal is often referred to as an *attack* in abstract network studies (Albert et al., 2000). The most common strategies used for network attacks are either random, where nodes and edges are randomly selected and removed from the network, or targeted, where nodes and edges are selected based on their structural/topological importance in the network (Gallos et al., 2005). The random attack strategy resembles some real-world disruption events like random equipment failure, operational faults, and accidents (Pant et al., 2016). The targeted attack strategy, to some extent, aims at capturing events like terrorist attacks (Huang et al., 2011) or some theoretic worst case scenarios.

While random disruptions and targeted attacks are useful simplifications, they do not completely cover the wide variety of possibilities that a real-world scenario may present. Many disruptive events, particularly weather-related events, may not fit into either. Although weather events feature stochasticity, they are not purely random because the climate system has deterministic dynamics that exhibit chaotic behaviours (Elsner and Tsonis, 1992). Additionally, weather-related disruptions certainly do not maliciously target any specific network components, making the targeted attack strategy unrealistically severe. Those failure initiation approaches are not capable of capturing the feature of weather-related disruptions.

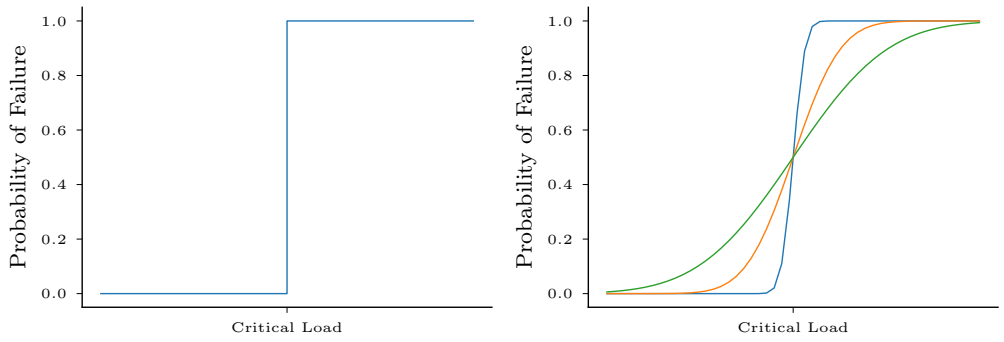
There exist extensive works in the literature on the resilience of infrastructure systems to extreme weather events (Bessani et al., 2018; Tari et al., 2021; Younesi et al., 2020; Trakas et al., 2019; Sabouhi et al., 2019; Panteli and Mancarella, 2015a,b; Panteli et al., 2016; Fu et al., 2018). The events are represented in various forms. For example, in Bessani et al. (2018), the weather condition across the power distribution system is characterized by two variables, the number of thunders and the maximum wind gust speed. This thesis is more interested in works that use weather profiles to initiate network disruptions and in particular, works that address the spatial profiles and patterns of weather events in their assessments, instead of using one single

homogeneous weather condition across the entire network. A common strategy is to divide the system under study into weather regions, within which the weather conditions are assumed to be homogeneous, e.g. Trakas et al. (2019) and Sabouhi et al. (2019).

Panteli and Mancarella (2015a) propose a conceptual framework to assess the influence of climate change on weather-related power interruption. In their work, an explicit reference is made to the use of weather profile data in both time and space domains to initiate system component failure and simulate cascading effects. However, subsequent works (Panteli and Mancarella (2015b) and Panteli et al. (2016)) appear not to implement such a strategy fully. The hourly wind profiles in Panteli and Mancarella (2015b) are obtained by sampling three normal probability distribution functions representing normal, high and extreme weather events. In Panteli et al. (2016), the space-varying aspect is achieved by dividing the network into 6 weather regions. Within each region, the weather conditions are assumed to be homogeneous and are sampled with a weather simulator. Works that account for actual spatial weather patterns, e.g., Fu et al. (2018), do so by reproducing historical wind extremes with spatial correlation. This implies the assumption that historical or present-day weather statistics will hold in the future, which is now widely recognized as a fallacy due to the changing climate and weather patterns (IPCC, 2013; Seneviratne et al., 2021).

Failure determination

Once a disruptive event is generated and the stress level on each network component is assigned, the next step is the determination of status for network components of interest: if the asset can withstand the assigned shock, and if not, to what degree the asset's service capacity is reduced.



(a) Step Function Form

(b) Sigmoid Function Forms

Figure 2.7: Example fragility curves. (a) A fragility curve in the form of a step function, which can be used to represent brittle and abrupt failures. (b) Three s-shaped functions representing probabilistic failure modes.

Previous studies have utilised a range of fragility functions to estimate if an asset can withstand the assigned stress level (Sutherland and Wolf, 2002; Panteli and Mancarella, 2015b; Panteli

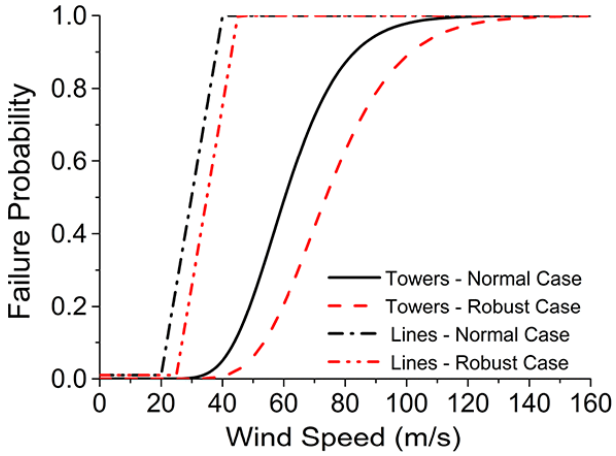


Figure 2.8: The wind fragility curves of transmission lines and towers for the normal and robust networks by Panteli and Mancarella (2015b)

et al., 2016; Fu et al., 2018; Camus et al., 2019; Sanchis et al., 2020). A fragility function expresses the probability of failure as a function of the magnitude of the hazard. The shape of a fragility curve reflects the uncertainty in the asset's ability to withstand a shock.

If the failure is deterministic, the fragility function takes the shape of a step function with a threshold and the probability of failure passes abruptly from 0 to 1 once the threshold is exceeded (Figure 2.7a). Examples used in the literature are air temperature threshold for railway track buckling (Sanchis et al., 2020) and significant wave height on port operation (Camus et al., 2019). In Sanchis et al. (2020), two forces are used to determine if a track buckles. The track buckling load (F_T) is calculated as a function of the difference between the temperature of the track and the temperature when the track is installed. The track buckling strength (F_B) is determined by track component and misalignment data. In their simulation, the track buckles if $F_T > F_B$. In Camus et al. (2019), a threshold of 0.4m for the significant wave height is used. Any hours where the simulated significant wave height exceeds 0.4m is deemed hours of non-operability. The threshold model, or step function, assumes that asset failures instantly and deterministic when the external stress level is above the specified threshold.

When there are greater uncertainties in the asset's capacity to withstand a shock, a more general sigmoid function can be used (Figure 2.7b). Examples include Panteli and Mancarella (2015b), Panteli et al. (2016), and Fu et al. (2018), where sigmoidal fragility functions are associated with the failure of electricity transmission lines and towers due to local wind speed (Figure 2.8). The fragility functions are a key component of the resilience assessment framework, but the construction of such fragility functions is usually not part of the resilience assessment. The construction and validation of such fragility functions are complicated and are separate research efforts, the research outcomes of which, became the inputs of infrastructure resilience assessments. The construction of the fragility functions are often carried out through empirical method (Bakalis and Vamvatsikos, 2018; Baker, 2015), analytical methods (Silva et al., 2019), expert judgements (Porter, 2015), and operational standards (Camus et al., 2019).

For a shocked asset, the reduction of its service capacity can be either complete or partial. A

binary function is the simplest type of function used, where there are only two outcomes: 0 as complete failure and 1 as normal functioning. Though the binary state function seems an over-simplification of actual asset response, it has the advantage of massive saving in computational power. Its application to infrastructure systems can be found when the paper either focuses on quantifying catastrophic consequences and possible worst-case scenarios, or attempts to evaluate interdependence across several infrastructure systems where the network studied is of great size, e.g., Pant et al. (2016). Partial reduction expands the discrete {0,1} binary state to a continuous space. Instead of only two outcomes, the service capacity reduction of impacted assets can take any value between 0 and 1. However, lack of empirical data limits its application. Magnitude of reduction are normally presumptive. In Zhang et al. (2015), a 50% reduction is randomly assigned to half of the shocked assets and 90% reduction to the other half. It is generally difficult to justify the viability and reliability of the speculated degree of reduction.

After initiating asset failures and determining the component-level service capacity reduction, failure propagation and cascading effects are considered in some work. Cascading failure refers to the chances that small local failures could spread globally and cause a systemic collapse due to inter- and intra-system dependencies. Those interactions are normally described as interdependences if the network studied involves multiple sub-systems standing for different infrastructure sectors and the interacted assets belongs to different sub-system. Rinaldi et al. (2001) identifies four types of infrastructure system interdependencies, namely physical, cyber, geographic, and logical interdependency, which underpins the foundations for infrastructure interdependencies studies. Since their work, a large amount of system interdependencies is identified and studies and has seen its integrations into resilience assessments (Pant et al., 2011; Chen and Miller-Hooks, 2012; Pant et al., 2016; Ouyang, 2017; Goldbeck et al., 2019; Smolyak et al., 2020; Sun et al., 2021). When those dependent effects are considered, propagated asset failures need to be calculated before estimating the system-level service disruption.

Asset recovery

As discussed before, the inclusion of recovery is the key difference between robustness and resilience. The recovery stage completes the resilience curve and allows the estimation of the total loss of services caused by the disruption event. Approaches used in reviewed literatures for asset recovery are presented below.

- **Importance-based method** is similar to some of the network attacking strategy, where important assets are subjected to higher priority to recover. Baroud et al. (2014) proposes two importance-based recovery strategies, where the importance criteria are set as magnitude of service loss cause by the impacted asset or time required to recover the asset. This approach, unlike the targeted network attack, has a slightly higher level of representability to real-world cases, as major infrastructure assets can be assumed to be more prepared for shocks and accessible to more resources.

- **Stochastic method** generally utilises the Monte Carlo simulation. In Aydin et al. (2018),

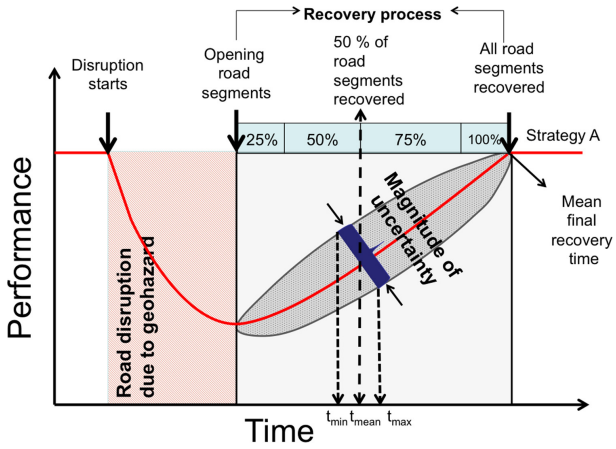


Figure 2.9: Conceptual resilience curve where the recovery phase is presented with uncertainty. Figure from (Aydin et al., 2018).

recovery time for each asset (road segment) is sampled through a Monte Carlo simulation with three different probability distribution functions. The resulted system resilience curve allows the estimation of the magnitude of uncertainty when there is no or limited information on each asset's recovery speed (Figure 2.9).

- **Generic presumption** is used for networks with a small number of assets, where the priority and recovery time for each asset can be explicitly specified. Chen and Miller-Hooks (2012) proposes six generically defined recovery scenarios. For each scenario, it details the applicable assets, recovery time, recovery cost, and percentage increase in asset capacity. Similar set of preparedness and recovery activity is used in Miller-Hooks et al. (2012). However, this method is only applicable to networks of small size. In Chen and Miller-Hooks (2012) and Miller-Hooks et al. (2012) the container shipping network used consists of 8 nodes and 12 edges.

- **Logic presumption**: In Aydin et al. (2018), in addition to the stochastically generated recover time, additional three road-segment recovery strategies are also adopted based on the following criteria: 1) proximity to the main resource centre; 2) proximity and hierarchy of roads; 3) Optimal proximity and recovery time required. Proximity refers to geographical distance and recovery time is estimated using surveyed data on the amount of debris on the road.

2.3.3 A Brief Summary

Overall, the concept of resilience provides a holistic view regarding a system's ability to withstand external disruptive events. A systematic resilience assessment framework can be developed for the assessment of climate resilience of infrastructure systems. However, an important limitation of current resilience assessment frameworks in relation to the effects of climate change on infrastructure systems is that, weather-induced failure scenarios are not simulated with future weather profiles but simulated with random removal, targeted removal, and manipulations of present-day weather statistics. A synthesis of summaries is provided at the end of this chapter.

2.4 Resource Allocation Strategies towards Climate Resilience

A changing climate continues to be a challenge for all infrastructure sectors (Betts et al., 2021). Early climate change adaptation investments can save money in the long term (JCNSS, 2022). AR6 defines climate change adaptation as *'the process of adjustment to actual or expected climate and its effects in order to moderate harm or take advantage of beneficial opportunities'* for human systems (as opposed to nature systems) and recognises the criticality of adaptation in minimising the impacts of climate change (Pörtner et al., 2022).

Recognising the importance of high-quality infrastructure in economic growth and the risks posed by climate change, the UK government has set out long-term plans for infrastructure investment across the nation (HM Treasury, 2020). As part of these plans, the government has made the economic commitment of investing 1.1% - 1.3% of the GDP each year from 2025 to 2055 to deliver the infrastructure development plans made (Treasury, 2021). In the United State, the government announced a 1$ trillion support for national infrastructure development, of which $47 billions are designated for infrastructure resilience, to prepare the nation for the changing climate (The White House, 2021). With resources made available to develop and upgrade the infrastructure systems, the question is how resources should be allocated most efficiently to minimise the forthcoming service disruptions. Two fields of research have addressed similar problems and developed approaches to answer similar questions as asked here.

Network science studies the topology and structure abstracted graph models, which inspires and drives the development of a large number of network metrics for identifying critical nodes and edges in the network. Those metrics provide ways to quantify the 'importance' of network components and have the potential to aid the prioritisation of asset upgrades. The second research field is at the intersection of optimisation and operational research. Methods developed in this field begin with constructing a mathematical or computational model that abstracts the essence of the infrastructure system. With the constructed model, the infrastructure system's health, safety, functionality, reliability or resilience is expressed as an objective function. With the resource input into the system acting as one of the constraints, the optimisation process then seeks the maximum or minimum of the objective function within the constrained feasible region, thus directly solving the problem of where to allocate the inputted resources.

2.4.1 Network Metrics as Importance Measures

Infrastructure systems are often modelled as networks. As pointed out in the sections before, infrastructure systems are complex systems where the behaviour of the system cannot be easily inferred by the behaviour of its components. The complexity comes from not only the variety and the number of components but also the interactions and relationships between those components. Network Science and graph theory provide a way to map out both the components and the pairwise connections. The abstraction of infrastructure systems to graphs and networks provides a powerful framework for analysing the infrastructure system's topological structure

and identifying important assets.

Infrastructure systems are susceptible to all types of failures. Motivated by these risks, a field of research has evolved that investigates the system's 'error and attack tolerance' (Albert et al., 2000) with respect to random failures and targeted attacks, using abstracted mathematical graph models. Building upon the concept of this error and attack tolerance, a diverse range of research fields have formed and led to the development of various network metrics related to the networks' robustness, vulnerability and resilience.

One of the basic features of a network metric is whether the metric is local or global. Global metrics typically describe network-level properties, capturing features of the entire network. Local metrics instead measure each individual node's or edge's property, their attribution to network robustness, or their liability to faults. As the resource allocation problem here concerns the quantification of the importance of network components to inform asset upgrade priority, the focus is on the local metrics instead of the global ones.

The Three Fundamental Network Metrics

Metrics related to the *degree, clustering,* and *centrality* are discussed first as they are recognised as the three most fundamental and mostly used local network metrics (Oehlers and Fabian, 2021).

The degree of a node is defined as the number of edges connected to that node (Barabási, 2013). For a directed network, it divides into in-degree, the number of edges arriving at the node, and out-degree, the number of edges departing from the node (Barabási, 2013). Node adjacency is one of the most straightforward and intuitive measures for component importance. The idea is that nodes with a higher degree, more connected edges, are structurally more crucial than nodes with fewer attached edges. Removing a high-degree node could remove a substantial proportion of edges in the network and is more likely to collapse the original network into two or more disconnected sub-networks.

The clustering coefficient was firstly introduced in Watts and Strogatz (1998) to describe *'the cliquishness of a typical neighbourhood'*. It can be applied to the whole graph and acts as a global metric. It can also be applied to a subgraph that is formed by a node and its neighbours and used as a local metric. For an undirected and unweighted graph $G(V, E)$, suppose that a node v_i is of degree k_i. These k_i number of its immediately connected nodes are called the *neighbour* of node v_i. The maximum possible number of edges that can exist between node v_i and its immediately connected nodes is $k_i(k_i - 1)/2$, which only happens when every neighbour of node v_i is connected to every other neighbour of node v_i. The clustering coefficient for this small community immediately around node v_i is defined as:

$$C_i = \frac{e(v_i)}{k_i(k_i - 1)/2} \tag{2.1}$$

where $e(v_i)$ is the number of edges that actually exist in this subgraph, and $k_i(k_i - 1)/2$ is the

maximum possible number of edges that can exist in this subgraph. In this way, the clustering coefficient describes the cliquishness of the small community formed by node v_i and its immediately connected nodes. Equations to calculate the clustering coefficient for weighted graph and directed graph can be found in Saramäki et al. (2007) and Fagiolo (2007).

The clustering coefficient describes the local cohesiveness, or connectedness, of a node. The idea originates in social network studies, where it is observed that common friends of a person tend to know each other. Under the context of transportation or communication networks, the clustering coefficient of a node reflects the likelihood of there exist alternative routes without extensive increase in path length if the node fails (Edwards et al., 2012). In this sense, regarding the resource allocation problem considered in this work, the immediate instinct is that a node of lower clustering coefficient should be of higher priority as it is less likely to find an alternative route around the node if it fails.

The betweenness centrality is a family of measures based on the shortest path between all nodal pairs in the network (Barabási, 2013). The idea is that a component in the network is in a central position if it falls on the shortest path between many node pairs (Freeman, 1977). The betweenness centrality of a node is the number of times that all-pair shortest path passes this node. The betweenness centrality of an edge follows a similar definition, defined as the number of times that all-pair shortest path passes this edge.

The betweenness centrality of a node/edge directly indicates how many origin-destination pairs would have their original shortest path disrupted if the node/edge fails. Edges/nodes with high betweenness centrality values indicate that there could exist a narrow channel between two communities or disparate parts of a network. Failure of such an edge/node can lead to a situation where no path exists in the network between some of the affected origin-destination pairs or even lead the network to become disconnected. The removal of such an edge/node is more likely to cause increments in path length between those disrupted original-destination pairs as the original shortest path is no longer available. For infrastructure networks, a sensible choice is to prioritise and protect those of a higher betweenness centrality as their failure could cause bigger scale service disruption than assets of lower betweenness centrality.

Modified metrics

The three metrics above are developed with abstract networks and rely solely on topological information to quantify the importance of nodes. While they are recognised as the three most fundamental network metrics , they may lack the applicability to real-world networks, as these systems may not share the same network dynamic as the abstract networks. There has been works focusing on adapting those metrics to study real-world systems, resulting in the development of modified and tailored versions of the three fundamental metrics.

Newman (2005) argues that in some real-world situations, the path from an origin node to a destination node does not always follow the shortest path - for example, the spread of a virus, the dissemination of information and subjective preference on route selection. A piece of

information can wander in the network until it reaches its destination, and its path is unlikely to be shortest. In this sense, they proposed a new betweenness metric. The random walk betweenness of a node, v_i, is the number of times a random walk between two other nodes passes through this node, averaged over all node pairs. The classical shortest-path based betweenness centrality in Freeman (1977) and this random-walk betweenness centrality, in essence, sit at the opposite end of a spectrum of certainty. The former strictly follows a predefined optimal path to its destination, and the latter is of complete randomness in terms of finding a path to its destination. Some real-world situations may sit in between the two ends. For example, in electric power transmission networks, the flows could occur at routes that are not the shortest possible path due to various operational rules (Zio and Piccinelli, 2010). It is, of course, challenging for network metrics to capture all the dynamics of a system, particularly when there is human interventions. Nevertheless, the use of network metric should at least seek to take the network's dynamic into account.

Each infrastructure system behaves differently, and some tailored network metrics were developed to reflect this difference. For example, Giustolisi et al. (2019) further tailored the original betweenness centrality for a water distribution network. The paper claimed they made three main modifications. First, the edge betweenness centrality was used instead of the node one. Second, they introduced weight to the edges in order to factor in the hydraulic characteristics of the pipes. The first two modifications are more like a variation of original the betweenness centrality measure instead of modifications. Third, the source nodes were represented with a star of multiple fictitious nodes. In this way, each source node was modelled as a hub of nodes in respect to the hydraulic system's behaviour. Their results suggested that the tailored metric led to a more practical domain analysis, which can capture some emerging hydraulic behaviour of their water distribution system.

Stand-alone use of the network-based approach could yield insufficient information with respect to the component's probability of failure posed by patterns of external shocks. Components associated with high consequences are not necessarily prone to faults, particularly for weather-related failure scenarios. One way to integrate the probability of component failure into the network-based approach is to utilise the concept of vulnerability or reliability - the probability of failure multiplied by the consequence of failure.

The probability of failure can be embedded into edge or node weight before carrying out network analysis. In Cadini et al. (2009), treating the reliability of each edge in the network as its edge weight, three classical centrality measures - degree centrality, closeness centrality, and betweenness centrality - were adjusted into three reliability centrality measures - reliability degree centrality, reliability closeness centrality, and reliability betweenness centrality. The latter three measures are the weighted version of the three classical centrality measures with the given edge reliability. They applied both the classical and modified centrality measures to the IEEE 14 bus transmission network , which is a simple approximation of the American Electric Power system. Their analysis showed that the modified centrality measures rank the nodes differently from the classical ones. The inclusion of the reliability measure managed to spot

some vulnerable components of high importance that otherwise would not be identified with pure topology analysis. The reliability aspect is a valuable addition to the purely topological centrality measures.

Multiple network metrics can also be combined into one single metric to reflect multiple aspects of component importance. Dunn and Wilkinson (2013) compared five network metrics' ability to predict the importance of nodes in a sample water distribution network. The five measures are 1) the original amount of flow through the node (OF), 2) the degree of the node (D), 3) the betweenness centrality of the node (BC), 4) OF*BC, and 5) OF*BC/D. Their simulation suggested that the last measure, which included three aspects of the network, was superior in indicating the importance of nodes. Their explanation was that, OF, as a physically-based measure, captures the importance of nodes for a given supply-demand scenario. Whereas D and BC, as graph theory-based metrics, indicate the topological importance of the nodes in the network. Their simulation results suggest that the last metric, OF*BC/D, which combines all of the perspectives, can highlight important nodes more accurately than the rest of the metrics.

Beyond the reliability-based measures, some metrics utilise the concept of *liability*. The criticality of network components in some approaches is evaluated with the consequence of removal. The importance of a network component is evaluated by quantifying the drop in the system's performance/functionality caused by its removal, e.g., Latora and Marchiori (2005). This has the same underlying logic as the degree-based and centrality-based importance ranking. The system's performance is associated with the the graph's connectedness or throughput and is measured with a global network metric, for example, network connectivity and the average shortest path. Global network metrics such as the network connectivity and the average shortest path are competent in capturing changes in the network's topology when edges/nodes are removed. However, when applied to real-world infrastructure systems, they may lack the ability to reflect the drop in the network's functional performance. It was found that the relationship between the number of asset failures and the amount of service loss is non-linear (Li et al., 2022). The intuitive is that changes in for example, the average shortest path of the graph, does no scale linearly with the magnitude of service degradation. The use of liability-based metric should be one that measures functional liability instead of topological liability.

A component's importance, or criticality, is determined by both its topological importance (centrality in the network) and functionalities offered (Scherb et al., 2017). In Fujita et al. (2018), the importance of an element of the network is set using an evaluation function,

$$v(i) = \lambda T(i) + (1 - \lambda)F(i) \tag{2.2}$$

where $T(i)$ and $F(i)$ represent the topological and functional importance of a component respectively, and $\lambda \in [0, 1]$ acts as a weighting factor between the two aspects. However, it is unlikely that the functionality importance of a component in a complex system is entirely unrelated to its topological importance. In some cases, for example, in a transportation network, the critical transportation hubs tend to be of a high number of connections. It may not be wise to add together the two separately evaluated topological and functional importance as a certain extent

of overlapping can be expected. A similar evaluation function is also observed in works studying the social network, e.g. (Gaeta et al., 2021). The problem is that the topological importance could partially contribute to the functional importance, and part of the functional importance can be attributed to the topological importance.

From the three classical metrics that are on a pure topological basis, it is obvious that topological measures on itself is not enough in solving resource allocation problems for various infrastructure systems. While those metrics tailored to specific infrastructure systems, integrating the topological importance and functional importance of a network can be difficult. Nevertheless, it is not wise to evaluate the two aspects separately and then add up. It hence drives the need to establish a clear mechanism or framework that explains how these local component-level importance metrics are related to the network-level resilience. Once such a framework is in place, it can be used to develop a more comprehensive metric that expresses the importance of network components as a function of their topological importance, functional importance, and climactic vulnerability.

2.4.2 Optimisation Problems for Networked Systems

The concept of optimisation is about *'finding the best possible solution by changing variables that can be controlled, often subject to constraints'* (Martins and Ning, 2022). The basic formulation of an optimisation problem is:

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{subject to} \quad & g(x) = 0 \\ & h(x) \leq 0 \end{aligned}$$

where $f(x)$ is the objective function, $g(x)$ is the equality constraints and $h(x)$ is the inequality constrains. The optimisation has an extensive application in operational research, engineering problems and economics. Applying mathematical optimisation to engineering problems requires two steps:

1. frame the optimisation problem;
2. solve the problem

This literature review focuses on the first step of formulating the optimisation problem and reviews existing optimisation models formulated for resource allocation problems in infrastructure networks. A suitable solution finding algorithm can only be located after the optimisation problem is formulated. A linear objective function with linear constraints is best to be solved with linear programming. A non-linear objective function with non-linear constraints would need more complicated solution finding algorithms such as the generic algorithm. Therefore,

techniques and algorithms used to solve the problem are noted here but are not the focus of this literature review, as this work does not aim to improve the computational efficiency or accuracy of any existing solution-finding algorithms. Based on the literature review here, a suitable solution finding will be allocated in Chapter 5 after the optimisation problem and objective function for this project is formulated.

The objectives observed in the literature typically are a system-level performance measure of interest. The objectives formulated can be categorised into the optimisation of system performance under normal operating conditions, referred to as ' operational optimisation', and under external pressure, referred as 'resilience optimisation'. The following two subsections first review works on operational optimisation and then resilience optimisation before reviewing the solution-finding methods used.

Operational Optimisation

Under normal operating conditions, user satisfaction is a commonly used objective. An example of using user satisfaction as the objective function is the Network Utilisation Maximisation (NUM) problem. Kelly et al. (1998) considers a network with a set J of links with finite capacities C_j and a set of users wanting to communicate with another user via a known route $r \in R$, where R is a set of routes. A binary matrix, $A = (A_{j,r}, j \in J, r \in R)$, depicts these link-route dependences, with $A_{j,r} = 1$ if link j lies on route r. Otherwise $A_{j,r} = 0$. When a capacity x_r is allocated to route r for a pair of user, this used capacity will then have utility $U_r(x_r)$. This utility is interpreted as user satisfactory. The resource allocation problem considers how to maximise the total utility of the network with a given set of capacity on links and states:

$$\max \sum_{r \in R} U_r(x_r)$$

subject to:

$$Ax \leq C$$

$$x \geq 0$$

In their formulation, user satisfaction is expressed as a function of the amount of information exchange capacity allocated. Though infrastructure user's satisfaction is often directly affected by the quality of service provided, the conversion from the quality of service provided (or the quantity of the service provided) to user satisfaction could be highly complicated (e.g. Hickish et al. (2019) and Hickish et al. (2022)).

Instead of user satisfaction, some optimisation models developed use the quality of services provided as the objective function. The exact form of the quality of service can vary depending on the nature and dynamic of the infrastructure. For example, electricity transmission and distribution networks are inherently dynamic systems that always aim to achieve a balanced state. Therefore, optimisation models considering normal operating conditions often do not

involve resources to be allocated but devices to be controlled - the decision whether a network component (e.g. transformers, capacitor banks, energy storage devices, generators, and switches) is on or off to maintain the balance of the network (Momoh, 2017). Objectives observed from reviewed literature include minimising the power loss in the distribution network (e.g. (Rao et al., 2012)), transmission loss (Ding et al., 2015), and voltage drop (Vita, 2017).

Unlike the power networks' balance-oriented targets, for railway networks, one of the essential measures for the quality of services is the percentage of trains which ran their entire planned journey calling at all scheduled stations and arriving at their terminating station within a specified timeframe. It is a measure that reflects both the punctuality and reliability of the railway services (Draper et al., 2017). Minimisation of the passenger waiting time is a popular objective in train schedule optimisation (Barrena et al., 2014; Yin et al., 2017; Högdahl et al., 2019). Each infrastructure sector has its own performance measures to quantify the service provided, and it is natural to use those performance measures as the objective function.

The quality of services that an infrastructure system provides can hugely depend on the quality of its assets. It is of great interest to infrastructure service providers and stakeholders to ensure the assets are maintained to meet the operational and safety standards with limited finance and resources. Optimisation models are developed to effectively allocate limited budgets and resources to maintain and improve the overall condition of the network. For instance, Chakroborty et al. (2012) studied the optimal resource allocation strategies to improve the health of a road network. A given budget is made available for multiple maintenance activities. Each maintenance activity, if implemented, mitigates multiple distress contributors (e.g. rutting and pothole). Each distress contributor is responsible for multiple aspects of a road section's health (e.g. structural conditions, safety-related conditions). The health of the road network is the sum of the health of constituent road sections multiplied by their importance in the network. The optimisation model maps out the complex relationships between various maintenance activities and the health of the road network and helps direct the resources to where they could have the maximum impact.

Optimisation models formulated in the field of operational research and infrastructure system studies is a enormous field. Objective functions formulated covers various aspect of infrastructure operation at various levels, such as the user satisfaction, quality of services, and quality of infrastructure assets as discussed above. As long as an operational objective can be expressed as an algebraic expression, it can be employed as an objective function in an optimisation model.

Resilience optimisation

As introduced in the previous section, the concept of network resilience has always been closely related to the concepts of network robustness and vulnerability, which have been widely studied in the field of network science and graph theory. A large number of graph metrics have been proposed to measure the robustness or resilience of the network (Oehlers and Fabian, 2021). The robustness metrics in the following paragraphs refer to those *'global'* metrics that describe

network-level properties, which differs from those local metrics reviewed in Section 2.4.1 for network component importance ranking. Those network-level robustness metrics have been directly used as the objective function.

For example, Zhang et al. (2017) considers the enhancement of network robustness through shielding critical edges. In their work, the robustness of the network is measured by its edge-connectivity, defined as the minimum number of edges whose removal disconnects the network. Shielded edges are assumed resistant to any type of attack and cannot be removed under any scenario. Through identifying and shielding critical edges, it would require more edges to be removed to disconnect the network, and therefore the network is considered more robust. They studied the problem of finding the optimal set of edges to shield to guarantee the edge-connectivity between given pairs of nodes under independent random link failure and correlated localised failures. Using network metric as the objective function and optimising towards the chosen metric implies that the system's resilience or robustness is solely associated with the single topological feature considered (e.g. the edge-connectivity).

The nature of the system under study is a leading factor in choosing which network metric to use. Distance-based networks, such as transport networks and computer networks where travel times, distance and shipping costs between node pairs are of interest, often use path-related metrics, such as the shortest path. For example, Cappanera and Scaparra (2011) considers the allocation of protective resources among components of a network to minimise the impact of malicious attacks on the length of the shortest path. Connectivity-based network metrics are used where the system disfavours isolation or disconnection of network components, such as communication networks or power networks.

Augusti et al. (1994) presented an interesting development from pure edge-connectivity. They studied the optimal resource allocation to minimise the seismic risk for highway networks. The origin-destination connectivity, combined with the failure probabilities of edges, evolved into a reliability-based objective function. The objective function is the reliability of the highway network, which is defined as the probability of maintaining a connection between a source and a destination node when an earthquake hits the network. Their work considered only one pair of origin and destination. Such a definition of reliability has been adopted by a large number of researchers since then and see its application in optimisation of infrastructure systems (Liu and Frangopol, 2005; Frangopol, 2011; Todinov, 2011; George-Williams and Patelli, 2016; Han and Frangopol, 2023).

The objective function in Fiondella et al. (2015) adopts the concepts of vulnerability. The objective function is calculated as a direct sum of a component-level vulnerability measure, which is a function of the traffic on the link, the probability of link failure, the cost of re-routing the traffic, and the link's defence level. Similarly, in McCarter et al. (2018), the network vulnerability is defined as the expected flow of commodities. The probability of survival of a link is expressed as a function of the strength of disruption to the link and the strength of the defence of the link. The expected flow of commodities in the network is a function of all link flows under all failure scenarios.

This thesis adopts the definition of resilience as the cumulative service loss from the time of disruption to the time of full recovery as discussed in Section 2.3.1 . With this definition, the concept of resilience has two primary dimensions - 1) robustness, describing the scale of onset disruption, and 2) recoverability, the recovery rate. From the perspective of proactive or reactive actions, resource allocation for resilience can be categorised into pre-disaster and post-disaster. Pre-disaster strategies look at allocating resources to reinforce the infrastructure system so that the onset impact of disruptive events can be minimised, reflecting on the robustness dimension of resilience. Examples includes Sun et al. (2021), Lobban et al. (2021), McCarter et al. (2018), and Fang and Zio (2019). Several studies focus on allocating resources after an extreme event to minimise the restoration cost or expedite the recovery speed. These studies are more on the recoverability dimension of resilience, e.g. Zhang et al. (2018).

Optimising towards one dimension will ultimately contribute to the resulting resilience of the system. In a recent study, Karakoc et al. (2023) look at the trade-off between investing in strengthening components (reducing vulnerability) versus investing in restorative ability (enhancing recoverability) in terms of maximising seismic resilience. Their results suggest that a pre-event investment only scenario is more efficient than a balanced pre-/post event investment and more efficient than a post-event investment only scenario. In the zero budget for the pre-event investment scenario, the disruption level at the onset of disruption is remarkably higher than in other investing scenarios. Even though the system is equipped with a high recovery rate, the initial onset disruption is so high that the system ends up with the least seismic resilience.

Solving the Optimisation Problems for Networked Systems

The characteristics and nature of the objective function largely determine what solution-finding techniques and algorithms are applicable. Linear objective functions can be solved with linear programming techniques such as the simplex method (Hillier and Lieberman, 2001). Solving optimisation problems with non-linear objective functions require more complicated solution finding algorithms.

When the objective function is non-linear, the convexity of the function is often viewed as the 'watershed' between analytically easy and difficult non-linear optimisation (Palomar and Chiang, 2006). A function is strictly concave if its first derivative function monotonically decreasing over its specified range. If plotted, the curve bowed downward like a cave (Figure 2.10). If a straight line is drawn between any two points over the range, the line should lies entirely under the curve.

Abstracted models with synthetic or mathematically perfect objective functions are more likely to be solved analytically, e.g., Kelly et al. (1998). As the model gets closer to real-world systems and the mathematical expressions get more complicated, an analytical solution (an exact solution in the form of a mathematical expression) obtained with pen and paper is not observed in any of the reviewed literatures.

The network utility problem formulated in Kelly et al. (1998) was implicitly solved under the as-

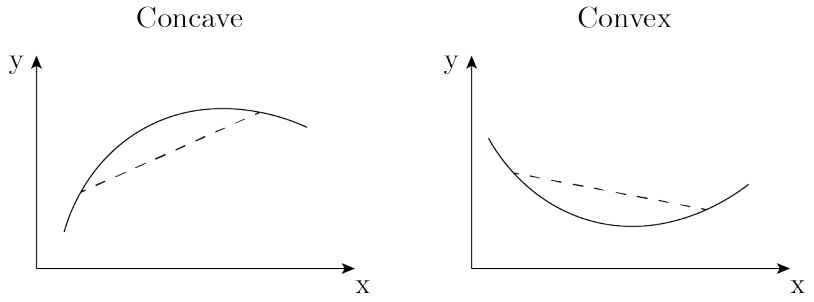


Figure 2.10: Concave and convex function

sumption that *'the utility $U_r(x_r)$ is an increasing, strictly concave and continuously differentiable function of x_r over the range $x \geq 0$'*. The original optimisation problem was decomposed into multiple, easier-to-solve sub-problems. The fascination of decomposition is that representing the original problem in a different way does not change the optimal solution but makes it easier to find the optimal solution. The problem is that not all optimisation problems are decomposable. The assumption of *strictly concave* and *continuously differentiable* makes the problem in Kelly et al. (1998) analytically solvable. However, objective functions from real-world systems rarely are strictly concave or decomposable.

Hande et al. (2007) recognise that there exist other types of traffic whose utility functions are not concave or continuous. They studied the additional conditions that need to hold if the utility function is sigmoidal or discontinuous in order to converge to the globally optimal capacity allocation. Their work suggests that when the assumption of strictly concave is removed, the optimisation problem becomes significantly more difficult to solve. Primarily because 1) it can not guarantee that the found local maxima is the global maxima, and 2) the difference in the values between the original objective function and the decomposed objective function can be always positive, meaning the solutions obtained from solving the decomposed problems can no longer guarantee the optimum.

Most of the infrastructure network optimisation problems share a certain degree of similarity with the Network Design Problems (NDP). Broadly speaking, NDP is about identifying the optimal selection of nodes and edges that satisfies some given constraints. For example, Johnson et al. (1978) presents a classical NDP, where given an undirected graph, they wish to find a selection of nodes and edges (effectively a subgraph) that 1) connects all the nodes in the original graph, 2) minimises the sum of all pair-wise weighted shortest path, and 3) the sum of its edge weight is within a given upper limit. Other well-known NDP includes the minimum spanning tree problem, the travelling salesman problem and the shortest path problem. The NDP framework provides a general and flexible representation for networked systems, and derivatives of the NDP have seen its applications in infrastructure systems, such as transport planning (Behbahani et al., 2019), post-disaster restoration (González et al., 2016), project selection problem for road networks (Hosseininasab and Shetab-Boushehri, 2015). González et al. (2016) introduces an interdependent network design problem (INDP) that takes system interdependence into account.

To a certain extent, the NDP is about finding the optimal set of network components for a

given purpose. The graph augmentation problem is similar in this way. It naturally implies a binary condition - a given network component can either be included in or excluded from the optimal set. Hence most of the optimisation problems mentioned above are solved with integer programming or other techniques for combinatorial problems. Fiondella et al. (2015) use the genetic algorithm to identify an effective allocation of a finite defence resource to 10 links in a rail network. The defence level of each link, the vector to be optimised, is assumed to take integer values between 0 and a pre-defined maximum defence level. The problem effectively becomes an integer programming problem. Even in the case of integral values only, the number of possible defence allocations among the 10 links is prohibitively large, making it impractical to compute all cases. Sun et al. (2021) studies a hypothetical collapse of a dam which would cause flooding over 13 bridges and affect the local transportation network and power plant. They considered the allocation of $3m budget to improve the absorptive capacity of the dam and the bridges. A commercial software package that combines multiple search engines (Tabu Search, Neural Networks, and Scatter Search) was used to solve this 10-dimensional optimisation problem (Kleijnen and Wan, 2007). Even a network model with 10 or 13 assets requires an overwhelming amount of computational powers. Any real-world infrastructure system would comprise of thousands of infrastructure assets and a representative infrastructure model would therefore forming an optimisation problem of over a thousand dimensions. The required amount of computational resources to solve such problem could be unrealistic to access.

McCarter et al. (2018) modelled the Swedish rail network of 1363 nodes and 1438 links as a directed network of 1363 nodes and 2876 edges. They aims to maximise the amount of flow under 5 sampled disruption scenarios where in each disruption event, 288 edges (10% of the total number of edges) were randomly chosen from the network and randomly assigned a "strength of disruption". A genetic algorithm with 50-member populations was used and 20 iterations were carried out to find the optimal allocation of defence resources. They did not report the time taken to run the simulation but used the phrase 'tens of hours' instead. When the size of the infrastructure network increases to a reasonable size, huge compromises are made in sample size, population size, and number of iterations, which can sacrifice quality of the solutions. On the positive side, those compromises made the problem computational tractable and therefore the algorithm is able to yield solutions to provide insights.

2.4.3 Brief Summary

From the perspective of priority ranking of infrastructure assets using network metrics, existing ranking / network-based methods appear to focus more on the topological structure of the network and overlook the supply-demand feature of the infrastructure systems. From the reviewed literature, there lacks a systematic way to integrate the network component's topological and functional importance into one single metric. Further, though there are works that bring the probability of failure into the metrics, to the best of our knowledge, no literature specifically uses climate model outputted weather data to set the objective functions. It hence drives the need to develop a more comprehensive metric that expresses the importance of network components

as a function of their topological importance, functional importance, and future weather-related failure probabilities.

From the perspective of mathematical optimisation, the objective functions are deceptively easy to formulate - as long as a concept, resilience, health, or safety of the infrastructure system, can be quantified as an algebraic expression, it can be employed as an objective function for a mathematical optimisation problem. However, the constructed optimisation problems are often troublesome to solve. Analytical solutions are only found under strict assumptions and lack applicability to real-world problems. Real-world problems are rarely solved analytically. Classical numerical optimisation techniques such as linear programming, non-linear programming or mixed integer programming have played a major role in solving these problems. The trend in recent literature has been to use evolutionary and stochastic optimisation techniques. Due to the *'curse of dimensionality'* (Bellman, 1957), the size of the solution space and the complexity of the problem increase exponentially with the number of variables. Large-scale infrastructure networks with many assets where each asset is a dimension make the problem high-dimensional and computationally intractable.

2.5 Research Gaps

2.5.1 A Systematic Climate Resilience Assessment Framework

As discussed in Section 2.2, a framework is needed to progress climate change impact assessment on infrastructure systems from component-level asset damage estimation to system-level functional loss estimation. Section 2.3 recognises that the concept of resilience provides a holistic view regarding a system's ability to withstand external disruptive events and can be further developed for the assessment of climate resilience of infrastructure systems.

However, the current resilience assessment frameworks, at their best, simulate weather-induced failure scenarios with manipulations of present-day weather statistics. Complex networks behave differently under different attack scenarios (Berche et al., 2009). Even if the infrastructure network of interest is shown to be resilient to random or targeted network attack strategies, it is not necessarily resilient to climate, or extreme weather event-based disruptions. Therefore, the resilience of infrastructure systems to weather-related events should be simulated with realistic weather profiles in addition to the random and targeted strategies.

What is resilient to the current patterns may easily be fragile to the future ones. Future weather profiles produced by climate models are a promising option to assess the resilience of infrastructure systems to future extreme weather events, moving beyond the widespread extremes of random and targeted attacks, unable to capture the features and threats of climate change. Given the amount of data made available from climate change researchers, an approach to simulate failure scenarios through realistic future weather profiles generated by climate models is hence needed.

In order to develop such a climate resilience assessment framework, the objectives are to:

- propose a method to initiate network disruption using climate model output;
- assemble a resilience assessment framework that is able to integrate such a failure initiation approach;
- benchmark the method through a case study.

2.5.2 A Resource Allocation Strategy towards Climate Resilience

Two complementary research approaches were identified as relevant to the problem of finding the optimal or best possible resource allocation between infrastructure assets to minimise the system's service disruptions under plausible future extreme weather events with a finite budget. The calculations of network metrics are of low computational complexity, but the network-metric-based ranking method lacks a clear mechanism to integrate the topological feature, the supply-demand feature and probabilistic failure patterns together. The mathematical optimisation method starts with defining an optimisation model that clearly maps out the relationships between the objective function and the variables. However, they often turn out to be complicated to solve and computationally intractable. The two methods have their own strengths and weaknesses. A *good* network metric that builds upon a formulated optimisation model may eliminate the need to solve those high-dimensional non-linear programming problems.

In order to develop a best possible resource allocation strategy towards climate resilience, the objectives are to:

- formulate the resource allocation between infrastructure assets to minimise potential future disruptions as an optimisation problem with defined objective function and constrains;
- using the optimisation model formulated, derive a network metric that captures the topology, service and failure probability of assets to indicate the priority of asset upgrade;
- find an analytically solution of the optimisation problem formulated;
- find linkages between the developed network metric and the solutions of the optimisation problem.

2.6 Chapter Conclusion

This chapter reviews literatures in the field of climate change impact assessment on infrastructure systems, infrastructure resilience assessments and resource allocation problems for infrastructure systems. Two research gaps were identified with objectives listed to bridging such gaps. The next chapter presents the work done in an effort to addressing the first research gap identified.

Chapter 3

Climate Resilience Assessment Framework

3.1 Chapter Introduction

This chapter aims to address the first research question of how the magnitude of future infrastructure service disruption due to climate change can be estimated with uncertainties. Resilience assessment has emerged as a powerful tool for analysing system responses to adverse events, including climatic hazards. [1]

In the previous chapter, Section 2.3.1, the concept of resilience was introduced and discussed alongside two closely related concepts, vulnerability and robustness. Section 2.3.2 reviewed existing resilience assessments and found that the assessment frameworks in the reviewed literature primarily simulate disruptive events and failure scenarios as random or targeted removal of system components, or simulated with manipulations of present-day weather statistics. Therefore, this chapter develops a novel method to simulate failure scenarios through realistic future weather profiles generated by climate models.

Section 3.2 presents the proposed failure initiation method as part of a systemic climate resilience assessment framework. The proposed framework initiates network disruptions based on geographical weather data outputted from climate models and uses the Monte Carlo simulation to estimate the variations in the outcomes of probabilistic asset failure.

Section 3.3 presents a case study carried out to quantify the resilience of Great Britain's railway passenger transport system to high-temperature-related track buckling under a climate change scenario. A 95-year horizon on the resilience of the railway system is drawn. The results reveal the non-linear responses of the railway system to the increasing temperature and show that models considering random asset failures overestimate the system's resilience.

[1] The current chapter contains materials that were previously prepared for the following: Li, Q., Punzo, G., Robson, C., Arbabi, H. and Mayfield, M., 2022. A Systematic Approach to Climate Resilience Assessment of Infrastructure Networks

3.2 A Systematic Climate Resilience Assessment Framework

The key elements of the proposed resilience assessment framework, namely the network model, the failure scenario generation, and the probabilistic resilience measure, are described in more detail later in this section. The proposed framework uses a combined topological and flow model for infrastructure system modelling to separate the modelling of probabilistic asset failure and estimation of system-level service disruption using deterministic service-asset dependencies. The failure scenario generation uses standard weather profiles obtained from the Earth System Grid Federation (ESGF), which holds the most extensive collection of observational, reanalysis, and simulation data for climate change research (Cinquini et al., 2014). Those weather profiles are available as individual time series of climate variables, such as wind, precipitation, and humidity, with different time resolutions and geographical ranges. It is worth mentioning that resilience in this work does consider the recovery stage, where rerouting and repairing activities are concerned, but not detailed in this method section as they could be interchangeable depending on the case study or the infrastructure sector being analysed. In the case study section (Section 3.3), a rerouting and a repairing algorithm are presented as illustrative examples. The high-level architecture of the proposed framework is shown in Figure 3.1.

3.2.1 Network Representation of Infrastructure Systems

In this work, the infrastructure system is modelled as a multilayer network, constituting a service layer and one or more asset layer. This work considers the ultimate purpose of an infrastructure system is to provide a service or services. A resilience measurement that relates to the system's ability to maintain the delivery of services when exposed to external shocks would be better than one that relates to the extent of damage/change of its topological structure. By separating the service and asset layer, it is possible to initiate failures in the asset layer but measure the scale of disruption with regard to the service layer. The successful delivery of a unit of service can sometimes rely on multiple interdependent infrastructure systems with components in each system affected by different types of climatic hazards. With such a separation, it is also possible to incorporate multiple asset layers into the assessment, which are exposed to multiple types of climate hazards. Failures can be initiated in different asset layers caused by different climate hazards simultaneously while increasing the flexibility of the modelling framework to adapt to different infrastructure systems.

Though it is possible to introduce multiple asset layers, the following formulation introduces only one asset layer and one service layer, which will be indicated by subscripts α and ϕ, respectively. Hence, for the asset layer, assets can be thought as a graph $\mathcal{G}_\alpha = \{\mathcal{V}_\alpha, \mathcal{E}_\alpha\}$ where $\mathcal{V}_\alpha$ and $\mathcal{E}_\alpha$ indicate the set of nodes and edges, respectively. The framework does not need to differentiate node assets and edge assets rigidly. For layer α, $\mathcal{A}$ indicates the set of assets in layer α, with elements $\mathcal{A}_i,\ i = 1, 2, \cdots, .|\mathcal{V}_\alpha| + |\mathcal{E}_\alpha|$, with $|\cdot|$ indicating the cardinality of a set. The equivalent definitions for the flow layer ϕ are omitted for brevity.

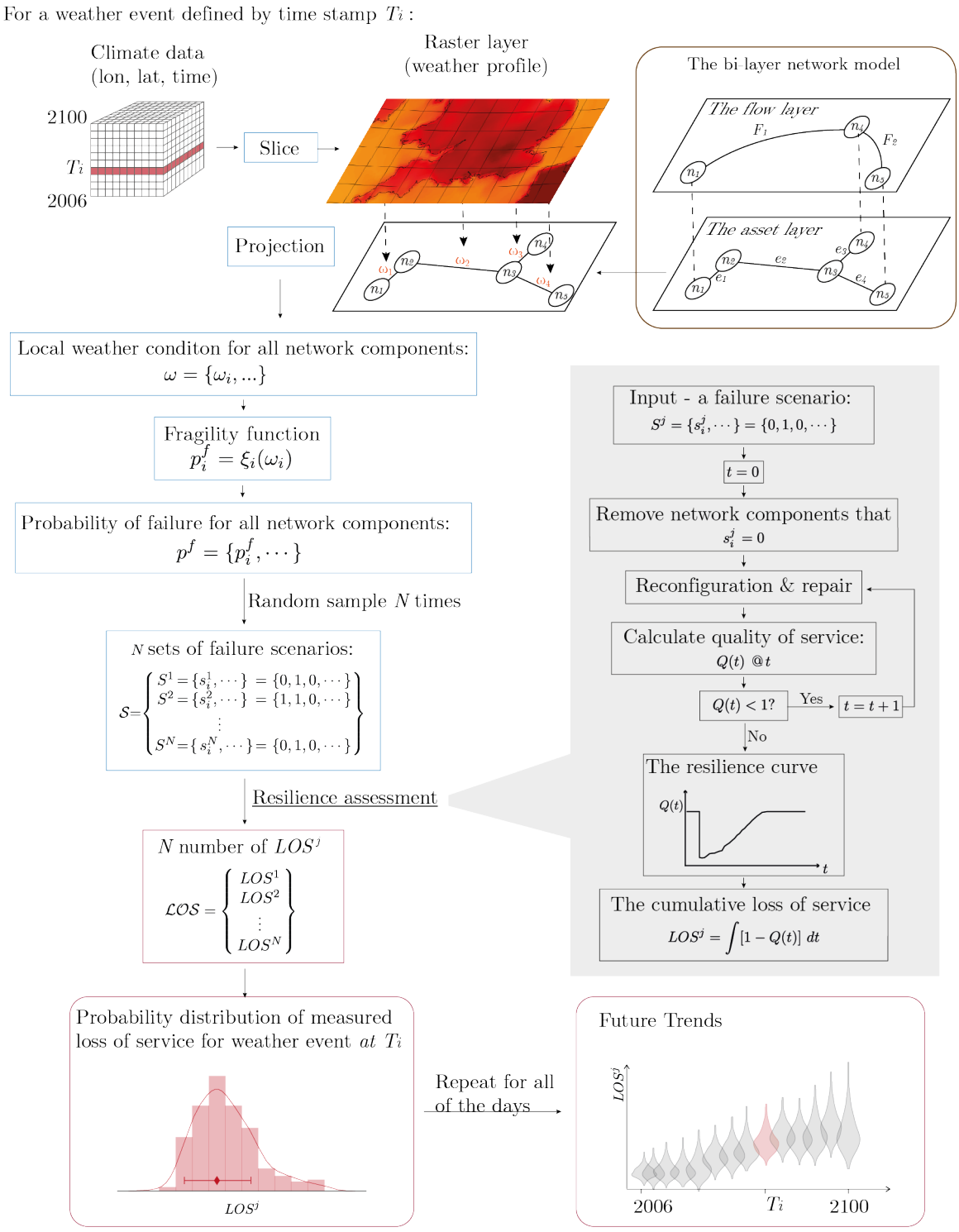


Figure 3.1: Overview of the probabilistic resilience assessment framework. Failure scenarios are sampled with weather profiles across the network. Service disruption is then calculated for each sampled failure scenario, forming a probability distribution. The resilience of infrastructure systems to future climate is assessed with a collection of such probability distributions.

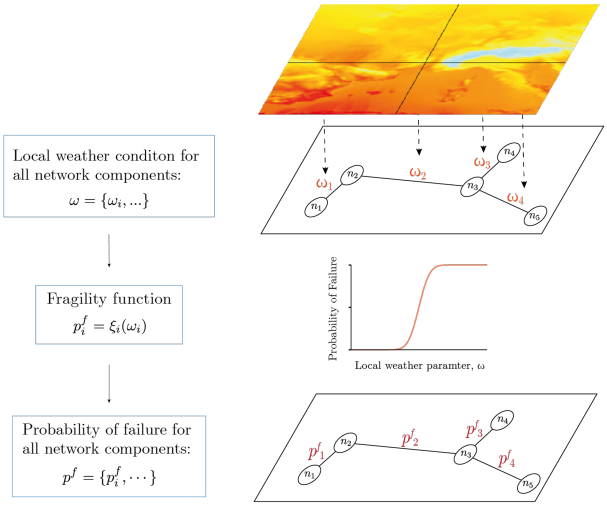


Figure 3.2: A illustration for projecting the weather profile on the asset layer and converting the local weather parameters to probabilities of failure.

3.2.2 Failure Scenario Generation

In this work, a *failure scenario* refers to a binary set consisting of the sampled asset failure, which is used to direct the removal of network components. The generation of such failure scenarios follows a systematic approach using geographically weather profiles.

As discussed before in the literature review, a weather hazard may feature a single or multiple weather parameter. To begin with, a climate hazard characterised by a single weather parameter is considered in this formulation. Climate models output single weather parameters as three-dimensional (3D) data, where spatial weather data are combined with the third dimension of time. A *weather event* refers to a slice from the 3D data with a desired geographical range for a given timestamp. As shown in Figure 3.2, for a given weather event, by projecting the geographical weather data on the geographical files of the asset layer, the local weather parameters of interest for all network components in the asset layer can be found:

$$\omega = \{\omega_i, ...\}, \ i = 1, 2, ...|\mathcal{V}_\alpha| + |\mathcal{E}_\alpha|, \tag{3.1}$$

where ω_i is the local weather parameter for asset $\mathcal{A}_i$.

With the local weather parameters allocated for each asset, an asset failure probability can be defined. This is the probability of an asset to fail due to the local weather parameter exceeding the asset's designed capability to withstand this types of weather hazard. This defines the likelihood that an asset can withstand an assigned value of a local weather parameter and can be derived from a fragility function. The fragility function returns a probability of failure as a function of the magnitude of the local weather parameter. In the most general form, consider the fragility function for each network component in the asset layer as:

$$p_i^f = \xi_i(\omega_i), \tag{3.2}$$

where ξ_i stands for the fragility function for asset $\mathcal{A}_i$ and p_i^f is the calculated probability of failure for asset $\mathcal{A}_i$ when the local weather parameter is ω_i. It is noted that the fragility function can vary from asset to asset depending on factors such as asset condition. The probability of failure for all network components in the asset layer can be written as:

$$p^f = \{p_i^f, \cdots\},\ i = 1, 2, ...|\mathcal{V}_\alpha| + |\mathcal{E}_\alpha|. \tag{3.3}$$

For the sake of simplicity, and without loss of generality, all assets are assumed to have a binary state. The state of a single network component, $\mathcal{A}_i$, can therefore be described by an independent Bernoulli random variable $X_i \sim Ber(p_i^f)$, so that

$$P(X_i) = \begin{cases} p_i^f, & \text{for } X_i = 0 \\ 1 - p_i^f, & \text{for } X_i = 1. \end{cases} \tag{3.4}$$

The state of the whole network is therefore effectively controlled by $|\mathcal{V}_\alpha| + |\mathcal{E}_\alpha|$ number of independent Bernoulli random variables:

$$\mathcal{X} = \{X_i, \cdots\},\ i = 1, 2, ...|\mathcal{V}_\alpha| + |\mathcal{E}_\alpha|, \tag{3.5}$$

with each corresponding to a failure probability.

Sampling each random variables once and the outcomes form a failure scenario, where asset failures are initiated by removing assets with $X_i = 0$ (Figure 3.3). Each failure scenario represents one possible outcome from the weather event. The complete sample space for $\mathcal{X}$ is of size $2^{|\mathcal{V}_\alpha|+|\mathcal{E}_\alpha|}$. To reduce the complexity of simulating the system's response to all $2^{|\mathcal{V}_\alpha|+|\mathcal{E}_\alpha|}$ combinations, a Monte Carlo approach is used to estimate the possible outcomes. Through repeated sampling, N set of possible failure scenarios can be obtained:

$$\mathcal{S} = \{S^j, \cdots\},\ j = 0, \cdots, N, \tag{3.6}$$

where j denotes a single Monte Carlo run. Each subset, S^j, contains a combination of states for all network components and is regarded as a single failure scenario.

$$S^j = \{s_i^j, \cdots\},\ i = 1, 2, ...|\mathcal{V}_\alpha| + |\mathcal{E}_\alpha| \tag{3.7}$$

3.2.3 Probabilistic Resilience Measure

This work considers resilience as the ability of a system to maintain and return to its normal operation after a disruption occurs. Here the resilience metric proposed by Bruneau et al. (2003) is adopted for to its general applicability. The resilience, R, is measured as cumulative service degradation from the time of earthquake, or any disruptive events, happening, t_0, to the time of full recovery, t_1, as shown in Equation (3.8). $Q(t)$ denotes the quality of infrastructure at time

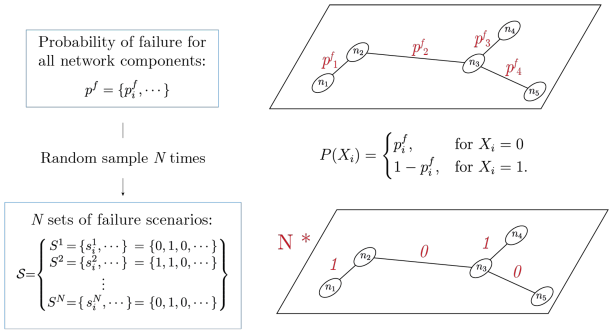


Figure 3.3: A illustration for the sampling process.

t in percentage, with $Q(t)$= 100% meaning fully functional status.

$$R = \int_{t_0}^{t_1} [100\% - Q(t)]\ dt \tag{3.8}$$

In this work, the resilience metric from Bruneau et al. (2003) is modified in two aspects. First, instead of the quality of the infrastructure assets, $Q(t)$ here measures the quality of service provided. Moreover, such a measure of the quality of services may take different meanings depending on the nature of the infrastructure system under study. As a general case, it is expressed as the percentage of satisfied demand:

$$Q(t) = \frac{\sum_\phi F^t}{\sum_\phi D^t}, \tag{3.9}$$

where F^t and D^t denote the delivery and demand at time t respectively.

The resilience metric is then calculated as cumulative loss of service, LOS.

$$LOS = \int_{t_0}^{t_1} 1 - \left(\frac{\sum_\phi F^t}{\sum_\phi D^t} \right) dt \tag{3.10}$$

Second, to take the uncertainty of the system behaviour into account, resilience to a given weather profile is associated with the averaged LOS from all sampled failure scenarios. This is achieved through a random sampling process. For a given weather event, a set of N failure scenarios, $\mathcal{S} = \{S^j, \cdots\},\ j = 0, \cdots, N$, are randomly sampled. For each failure scenario, S^j, component failures are initiated in the asset layer accordingly and disruptions are then computed regarding flows in the service layer. Reconfiguration and repair activities can be then performed. $Q(t)$, is calculated at every time step until full service is recovered. The cumulative loss of service, LOS_j, from the j-th sampled failure scenario, S^j, is then calculated with Equation 3.10. The statistical distribution of set, $\mathcal{LOS} = \{LOS_j\},\ j = 0, \cdots, N$, describes the system's resilience to the given weather event.

3.3 Resilience of The UK Rail Passenger Network to High Temperatures

Heat-related track buckling is one of the most common reasons for delays and cancellations in railway services. The UK experienced a brief but unprecedented extreme heatwave from 16 to 19 July 2022, where rail services were severely disrupted (Met Office, 2022). Recognising the urgency and consequence of this issue, this section presents a case study on Great Britain's railway passenger transport system, which is modelled as a flow network dependent on a single asset layer of train tracks subjected to high-temperature-related track buckling. A set of temperature projections covering the years 2006-2100 under a climate change scenario is used to generate plausible climate-based failure scenarios. Those failure scenarios are then used to initiate failures in the network, followed by a resilience assessment.

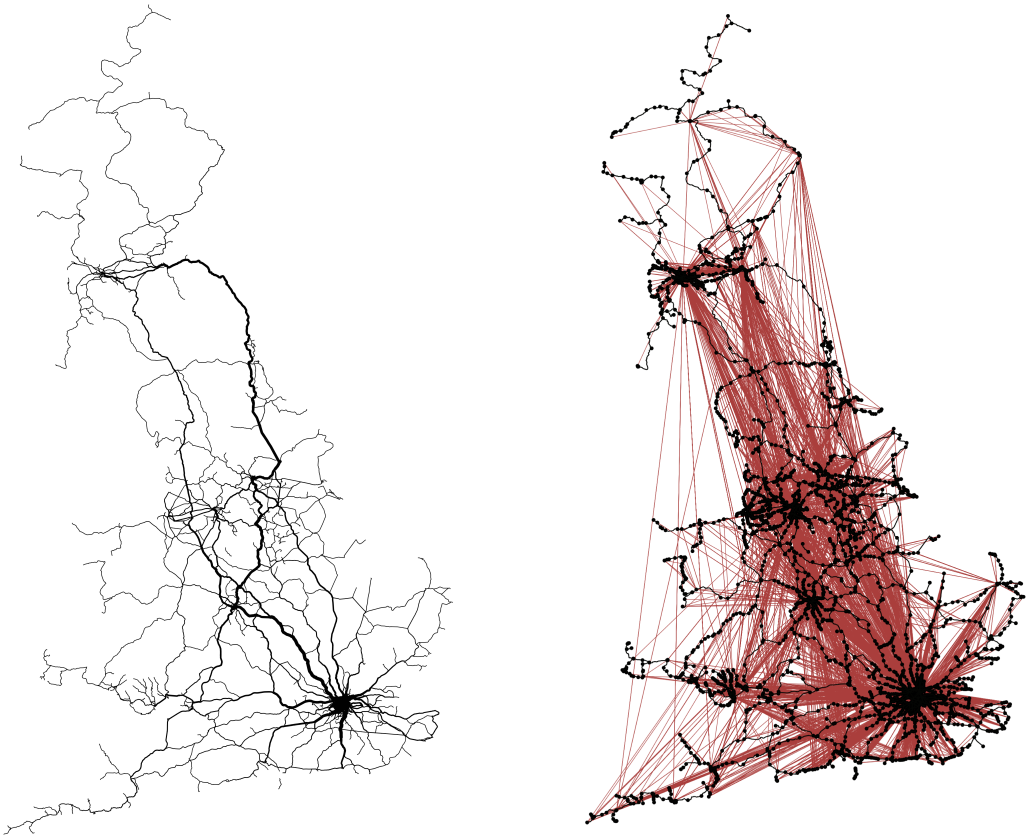
3.3.1 Model Inputs and Assumptions

Before presenting the case study, it is worth noting that this case study is more of a prototype for the climate resilience assessment framework. To complete the prototype, it needs a working network model, a specified climate hazard, a set of climate model output data, a rerouting algorithm, and a recovery strategy. In the absence of relevant information, several simplifications and assumptions are made in the most general and possible form to complete the prototype for the assessment framework. The assumptions made are detailed in the following subsections.

The Network

The railway network model developed by Pant et al. (2016) is used in this case study. It has separated flow and asset layers (Figure 3.4). The asset layer is an undirected weighted network that consists of 4024 stations, modelled as nodes; and 4524 railway track segments, modelled as edges. The edges in the flow layer is in the form of OD trips, representing the services provided. There are 2,282,270 OD pairs between 2484 origin and destination nodes out of the 4524 nodes in the asset layer. Each OD pair has an original edge path assigned, $\mathbb{P}^0_{od}$, detailing which edges in the asset layer are utilized.

The flows of the service layer are in the unit of annual average daily passenger trips (AADPT), which are of a temporal resolution of days. Therefore, this resilience assessment case study adopts the same temporal resolution and regards a day as a time step in the simulation. The average annual daily traffic is used as both a measure of the traffic volume on edges in the asset layer and OD demand. The system is assumed to have the same amount of services demanded every day, regardless of the state of the asset layer. Any daily, weekly, or seasonal variations are not considered in this case study.



(a) (b)

Figure 3.4: (a) Asset layer of the network model with 4024 nodes and 4524 edges. Edges weights indicate the amount of flows on edge with a minimum of 0.2 pixel. (b) Flow layer of the network model. This figure only plots 22,326 out of the total 2,282,270 OD pairs, which has more than 15 passenger trips and geographical path length greater than 30km.

The Hazard

High-temperature-related track buckling is used as an example hazard to demonstrate how to generate network failure scenarios using weather profiles. Railway tracks are made of steel, which expands under with increasing temperature. When the lateral force caused by the expanded track exceeds the resistance provided by the fastening, sleeper and ballasts, the track buckles. It is established that track buckling is related to the daily maximum temperature (Palin et al., 2013; Ferranti et al., 2016). To the best of our knowledge, there is no existing fragility function that expresses the probability of track buckling as a function of the ambient temperature. The empirical definition of such a function is beyond the scope of this work.

A Gaussian sigmoid is hence used as a fragility function. The rationale is hereby explained. To begin with, assume that a physical asset was constructed perfectly to its design load, μ_i, and behaves rigidly. Under this assumption, an asset functions if the local weather parameter is below the design load, $\omega_i \leq \mu_i$ and fails immediately if the design load is exceeded. Realistically, the local weather parameter that an asset can actually handle (the true failure load) could be higher or lower than the design load, taking into account the application of the safety factor, ageing,

and maintenance of the assets. If the sample size is large enough, the empirical distribution of the true failure load can be described by a random variable $\mathcal{X}$ that is approximately normally distributed, $\mathcal{X} \sim \mathcal{N}(\mu_i, \sigma_i)$, where μ_i denotes the design load that all the assets are supposed to have, and σ_i expresses the average difference between the design loads and the actual failure loads. For a given external weather parameter, ω_i, the probability of failure for edge E_i equals the probability of edge E_i's actual failure load smaller than ω_i.

$$p_i^f = P(\mathcal{X} \leq \omega_i) \tag{3.11}$$

The cumulative density distribution function (CDF) of the standard normal distribution is[2]:

$$P(X \leq x) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]. \tag{3.12}$$

Therefore, the probability of failure for edge E_i under external weather parameter ω_i is:

$$p_i^f = P(\mathcal{X} \leq \omega_i) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{\omega_i-\mu_i}{\sigma_i\sqrt{2}}\right)\right]. \tag{3.13}$$

The fragility function, ξ_i, for asset $\mathcal{A}_i$, therefore expresses as

$$p_i^f = \frac{1}{2}\left[1 + \text{erf}\left(\frac{\omega_i-\mu_i}{\sigma_i\sqrt{2}}\right)\right], \tag{3.14}$$

which is controlled by 2 parameters σ_i and μ_i. μ_i is related to the Critical Rail Temperature (CRT), which is dependent on other factors such as the stress-free temperature of the rail and the degree of consolidation of the ballast (Network-Rail, b), and σ_i is related to the confidence/uncertainty of the estimated CRT.

The critical air temperature for heat-related track buckling varies depending on track conditions. Therefore, ideally, the fragility function's shape control parameters, σ_i and μ_i, should differ from track segment to track segment depending on track conditions. However, track conditions are not available from the network model obtained or from other sources. It is therefore assumed all edges in the network are subjected to the same fragility function of the same σ_i and μ_i. Taking the thresholds in Palin et al. (2013) and Vajda et al. (2014) into consideration, the cumulative distribution function of $\mathcal{N}(35, 2.5)$ is used and all edges are subjected to the same fragility function that expresses as:

$$p_i^f = \xi(\omega_i) = \frac{1}{2}[1 + \text{erf}(\frac{\omega_i - 35}{2.5\sqrt{2}})], \tag{3.15}$$

where p_i^f is the probability of failure for edge i when the external air temperature is ω_i. The values of μ and σ can be tuned if more information becomes available, e.g., a series of historical failure events and corresponding local temperatures.

[2]erf denotes the error function $\text{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$

The Climate Data

As mentioned above, track buckling is related to the daily maximum temperature and the resilience assessment here regard a day as a time step because the flows in the flow layer are annual average daily traffic. Therefore the climate variable, daily maximum near-surface air temperature, *tasmax*, is used in this case study. EURO-CORDEX provides downscaled climate change data for the European domain, covering a period of 95 years, from 2006 to 2100. EURO-CORDEX provides data in two spatial resolution, 0.44°(∼50km) and 0.11°(∼12.5km). Between all climate scenarios, RCP8.5 is used in this case study in order to understand the worst possible scenario the railway system might face. By limiting the climate model output data to:

- 'variables' = 'tasmax',
- 'time frequency' = 'day',
- 'domain' = 'EUR-11' (0.11°),
- 'experiment' = 'RCP8.5',

64 sets of model output are left. The main differences between them are the global climate model used for downscaling, the climate model ensemble, and the regional climate model used. In the absence of a clear rationale to prefer one to the other, the most recently updated one when this work is carried out, EUR-11_CNRM-CERFACS-CNRM-CM5_rcp85_r1i1p1_MOHC-HadREM3-GA7-05_v2, is used in this case study. The effective climate sensitivity (ECS) of the climate model used is 3.25 whereas the average for 17 CMIP5 models studies in Zelinka et al. (2020) is 3.3, indicating that the climate model used, is not close to the 'extremes' in projecting future climate. It is noted here that, when a single climate model is used, the resilience assessment here can provide a sense of how broad the outcomes would fall and the associated likelihood of those outcomes for the climate model used. If computational power resources allows, more resilience assessment runs with multiple outputs from different climate models would give a more complete projection, which could show the degree of uncertainty and error range associated with climate models themselves.

Some preliminary testing runs suggest that the simulation time for a single weather profile increases with the national average temperature. This is because more edge failures can be expected for a hotter day and it takes more time steps to recover. In each step during the recovery stage, there are more than 100,000 search for a shortest path between OD pairs where the flows were disrupted in order to find paths for rerouting. Even with the most efficient shortest path search algorithm available, for some extremely hot days in the model, it would take up to 50 hours to complete the simulation for a single daily weather profile with 250 realisation in the Monte Carlo simulation. Therefore, a day-by-day resilience assessment for the 95-year period would hardly meet practical considerations and require a substantial amount of computational power.

In this case, a clustering analysis was carried out to select example summer days as high temperature are likely to happen in summer. Summer days (May to September) in each 5-year period are clustered together using a K-means algorithm where the optimal number of clusters (K) is obtained via an elbow analysis. The analysis is carried out with respect to the inertia (sum of squared euclidean distances of each clustered point to their cluster's centroid) of the clusters formed. By plotting the inertia against the number of clusters, the identified elbow is the point where the rate of decrease in the inertia slows down and adding more clusters than the elbow point will not significantly improve the quality of the clustering. The elbow analyses return results between 8 and 15 clusters. Therefore, a K-means algorithm is used to select 15 example summers days for each 5-year period. Practically, this is done using a Bisecting K-means algorithm, coded in the Python Scikit-learn package (Pedregosa et al., 2011), where all other parameters are set to their default values. The day closest to the synthetic centroid of each cluster is chosen as the representative day. By doing so, 285 example days are selected, representing 285 typical daily temperature profiles and thus 285 unique distributions of climate hazards over the railway network. This approach makes the dataset computationally tractable with the method presented without losing significant information and the 5-year interval is consistent with the size of the EURO-CORDEX data batches.

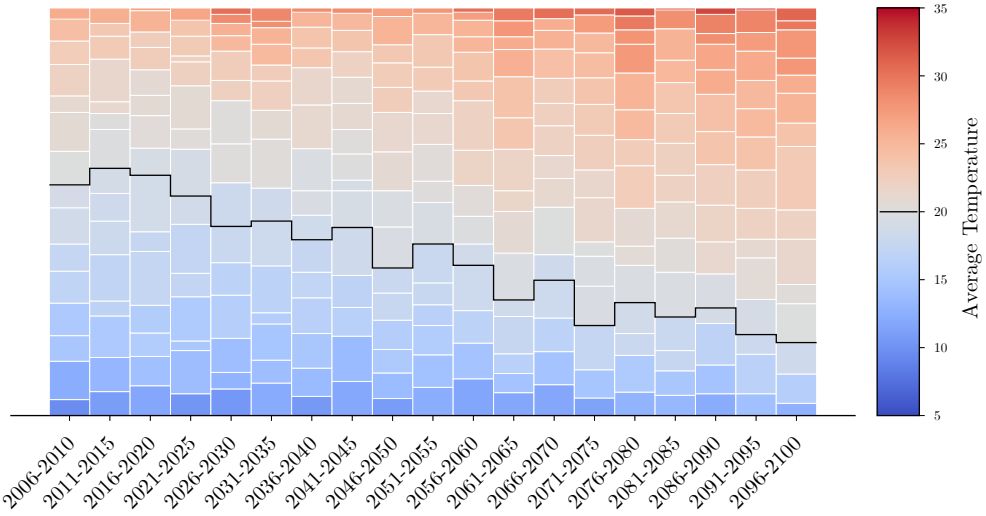


Figure 3.5: Clustering of the summer days (01 May to 30 September) from 2006 to 2100 based on the tasmax of EURO-CORDEX output data EUR-11_CNRM-CERFACS-CNRM-CM5_rcp85_r1i1p1_MOHC-HadREM3-GA7-05_v2. 15 clusters are produced for each 5-year group through a K-means algorithm. The height of each bar is controlled by the number of days that fall into the cluster. The color of the bar is controlled by the average temperature of the synthetic centroid of the cluster.

Rerouting Algorithm

Railways usually are not delivering at their maximum capacity (Ramdas and Chapman, 2017). When the network is partially damaged with some OD pairs losing their original path, the spare

capacity may in theory allow for rerouting as passengers could use the remaining running rail services to reach their destination by getting on a different train.

As no readily available and directly applicable traffic assignment algorithm can be found, a modified minimum-cost maximum-flow algorithm based on the Edmonds-Karp algorithm (Edmonds and Karp, 1972) is used for the flow assignment here (See ALGORITHM 1). Considering the size of the OD matrix and the time taken to search for a path in a network of great size, the algorithm is only set to search for the first 5 shortest paths instead of until no path exists. Further, the 'cost' here is the geographical length of the path instead of the number of steps (edges traversed).

ALGORITHM 1 - The Rerouting Algorithm

INPUT:

- R - residual functioning network structure.
- $\mathbf{c}_{u,v}$, - edge capacity
- $\mathbf{D}^{re}$ - OD flows to be rerouted.

$count = 0$

WHILE $count <$ Nr of path search:

- For all OD pairs that $D^{re}_{od} > 0$:
  - If there is no path...OK
  - If there is a path, P_{od}:
    - if path length $< 2 \times$ the original path length, DO:
      - temporarily allocated flow $T'^{re}_{od} = \min(c_{u,v},\ (u,v) \in P_{od})$
      - for all edges in the path: $f^r_{u,v}$ += T'^{re}_{od}
- For edge (u,v) in R:
  - If $f^r_{u,v} > c_{u,v}$:
    - reduction factor, $\varphi_{u,v} = c_{u,v}/f^r_{u,v}$
    - $c_{u,v} = 0$
  - Else $f^r_{u,v} \leq c_{u,v}$:
    - reduction factor, $\varphi_{u,v} = 1$
    - $c_{u,v}$ −= $f^r_{u,v}$
- For (o,d) in $\mathbf{D}^{re}$:
  - $\varphi_{od} = \min(\varphi_{u,v},\ (u,v) \in P_{od})$
  - T^{re}_{od} += $T'^{re}_{od} * \varphi_{od}$
  - D^{re}_{od} −= $T'^{re}_{od} * \varphi_{od}$
- Remove edges that $c_{u,v} = 0$ from R
- $count$ += 1

RETURN T^{re} - Trips delivered through rerouting

When the original path, $\mathbb{P}^0_{od} = \{e_{ij}, \cdots\}$ is interrupted due to one or more edge failures, passen-

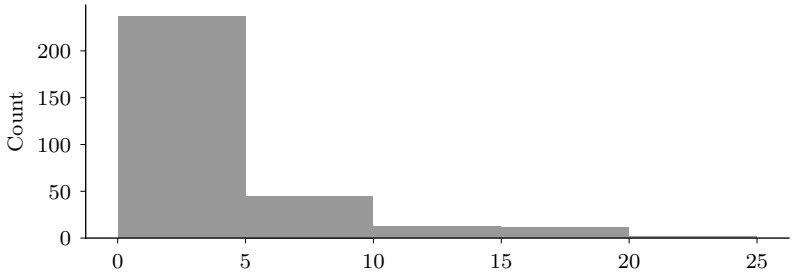


Figure 3.6: Histogram of the percentage difference between the amount of flow delivered at each time step between a full rerouting effort and the reduced rerouting effort. Twelve failure scenarios are included in the sample. The number of trips delivered with the two rerouting efforts is compared at every time step.

gers are assumed to stay at its original node v_o and no passenger is waiting at any intermediate nodes on the original path. Any edge, if undamaged, is assumed to have a spare capacity of 50% of its regular traffic for rerouting. At each time step, the algorithm implements 5 rounds of path searching. At each round of path searching, it searches for the shortest path that, 1) all edges forming the path has available spare capacity; and 2) is within twice the geographical length of the original path for each disrupted OD pair. The available capacity of the path is the smallest spare capacity across all the edges forming the path found (like a bottleneck). Any used spare capacity is deduced before the next round of path searching. The capacity is gradually used and in some cases, the capacities of some popular edges are reduced to 0 in the first round of path searching.

The algorithm involves a large number of path searches as there are 2,282,270 OD pairs in the flow layer with an asset layer constituting 2484 nodes and 4524 edges. The search for alternative paths is time-consuming and memory-intensive. To meet the limitation in computational power available, the algorithm is set to eliminate the alternative path search for any OD pair with fewer than 15 passenger trips or geographical path length less than 30km. Those trips can be regarded as trips that are likely to be aborted if there is a disruption in the railway due to their relatively small demand figures or met by using alternative transportation services due to their short geographical distance. These OD pairs contribute to 99% toward OD pair count and therefore consumes the majority of computation when searching for paths in the asset layer. By eliminating those OD pairs from the path search, the computational time is reduced by 96%. The sum of those OD pairs contributes 71% toward total passenger trips. The non-rerouted demand is assumed to resume travel as soon as the service is re-established on the original path. Up to that time, it counts as unsatisfied demand. A small-scale trial computation was carried out to assess the suitability of this approach. The difference in the absolute amount of flow delivered between a full rerouting and the reduced rerouting strategy chosen is less than 5% in the vast majority of the cases (Figure 3.6).

Recovery

As also discussed in the literature review, there are various assumptions and algorithms that can be applied to the repair process. This case study adopts the 0-assumption to keep the problem 'as simple as possible but not simpler' in order to keep the problem computationally tractable. Asset repairing activities are assumed to take place at every time step until the asset layer is fully recovered. For the sake of this example and to bypass the scarcity of information on the recovery of incidents in the railway sector, damaged edges are subject to the same recovery probability of 0.5 at every time step until the asset layer is fully recovered. Once an asset is recovered, no secondary damage is considered in this disruptive event.

3.3.2 Simulation

As discussed above, the flows of the service layer are in the unit of AADPT and the temporal resolution for the acquired climate variable *tasmax* is daily, a single day is regarded as a weather event. For a single day, the weather profile across the country is seen as an instance of situations the railway system might face. This case study considers continuous hot days or heatwaves that last longer than 24h as a single extreme event with repair work only happening after the event. This is consistent with Network Rail's operational standards, for which track maintenance works should not be carried out when the rail temperatures are above 32 °C (corresponding to an air temperature of 21 °C) or are predicted to exceed 38 °C (corresponding to an air temperature of 25 °C) within three days of work conducted (Network-Rail, a). For each weather profile, the local weather parameter for each edge are first assigned and transferred to the probability of failure using the fragility function. Then 250 sets of failure scenarios are generated through random sampling. The reasons why 250 sets of failure scenarios are chosen in terms of hardware, software and statistical data rationality are presented in Section 3.3.6

For a failure scenario, disruptions are initiated by removing edges for which $s_i^j = 0$ at time $t = 0$. Interrupted OD pairs, whose original path $\mathbb{P}_{od}^0$ includes any removed edges, are then identified. The rerouting algorithm then calculates the amount of OD flows rerouted. The total amount of delivery at this time step is the sum of the rerouted and the uninterrupted OD trips. The quality of service, $Q(t)$, is calculated with Equation 3.9. Undelivered OD flows, partially or completely, at the current time step, plus the steady-state daily demand (the annual average daily traffic) becomes the demand for the next time step. Asset repair is carried out between time steps until all edges are recovered. If the original path of an OD pair is recovered at a time step, from the next time step, any accumulated undelivered trips will be delivered via its original path with the spare capacity fully utilised instead of rerouting.

To compare the climate-based failure scenarios to random and targeted, shocks of the same intensity should be applied to the network in two separate settings with either a random or targeted strategy. This 'intensity' is taken as the number of edges removed in this work. As defined in Section 3.2.2, the state of a single edge is described by an independent Bernoulli

random variable $X_i \sim Ber(p_i^f)$, that:

$$P(X_i) = \begin{cases} p_i^f, & \text{for } X_i = 0 \\ 1 - p_i^f, & \text{for } X_i = 1. \end{cases} \tag{3.16}$$

If sampled $X_i = 0$, edge e_i will be removed and the probability for $X_i = 0$ is p_i^f. For a given weather profile, the expected number of failed edges, Ψ, for the whole network therefore can be calculated as the sum of the p_i^f for all edges:

$$\Psi = \sum_{i=1}^{|\mathcal{E}_\alpha|} p_i^f. \tag{3.17}$$

This implies that, for this given weather profile, on average, Ψ edges is expected to be removed in each Monte Carlo run. Therefore, for the random strategy, Ψ number of edges is randomly removed in each Monte Carlo run. For the targeted strategy, edges with the most through flows are removed first. In both cases, the same resilience assessment procedure as in the climate-based failure scenarios is followed.

Using Python, the simulations are carried out by Intel Xeon E5-2630 v3 CPUs. Each example day with a given failure strategy (climate-based, random, or targeted) is submitted as a batch job. Each batch job is allocated 16G of RAM. The CPU time for the 285 example days with climate-based failure scenarios is 1792 hours in total. The CPU time for random and targeted strategies is 1904 hours and 1881 hours respectively. For the experiments in this work, though the CPU time is around 1800 hours, a wall clock time of around 40 hours is achieved by submitting each example day's simulation as an individual batch job to the university's high-performance computing (HPC) services.

3.3.3 The 95-year Trend

For a given failure scenario, upon the edge removals at $t = 0$, the quality of service drops immediately. With rerouting and repairing efforts, the quality of service gradually bounces back

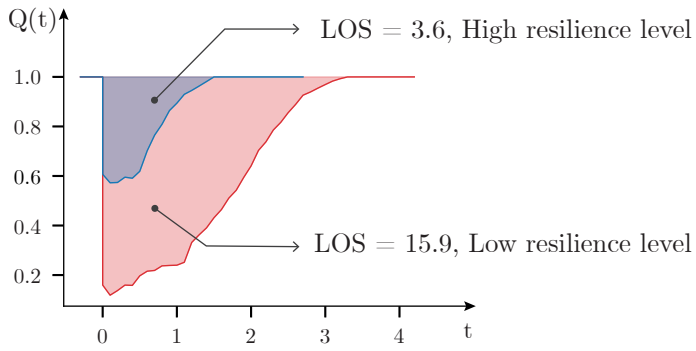


Figure 3.7: Illustrative resilience curve with calculated LOS as the area of the shaded 'triangles'.

to the fully functional level$Q(t) = 1$. This forms a system response curve. This curve and the horizontal line of $Q(t) = 1$ forms a triangle (technically not a triangle as the curve is not straight), the area of which, effectively is the cumulative loss of service and indicates how resilient the system is (see Figure 3.7). A high LOS value means a high degree of service loss overall and is therefore associated with low resilience. For a given weather profile, 250 failure scenarios are sampled and each failure scenario corresponds to a resilience curve with a calculated LOS. The distribution of the 250 calculated LOS shows the likelihood of the outcomes (loss of service) and indicates the system's resilience.

Figure 3.8 shows a time series of annual total LOS is constructed using the simulation results of the example days and the clustering analyses. This is done by duplicating the simulation results for each example day and assigning them to all days in the same cluster. The days are then grouped by year and the convolution of all non-zero distribution is computed to give the estimation of the annual total LOS. From Figure 3.8, the constructed time series data shows an overall upward trend regarding the estimated annual total LOS. Moreover, the spikes suggest that, in the climate model output used, there exist some extreme hot years with significantly more and hotter days. The overall upward trend suggests that the railway system's resilience to high-temperature-induced disruptions could be compromised under future climate scenarios.

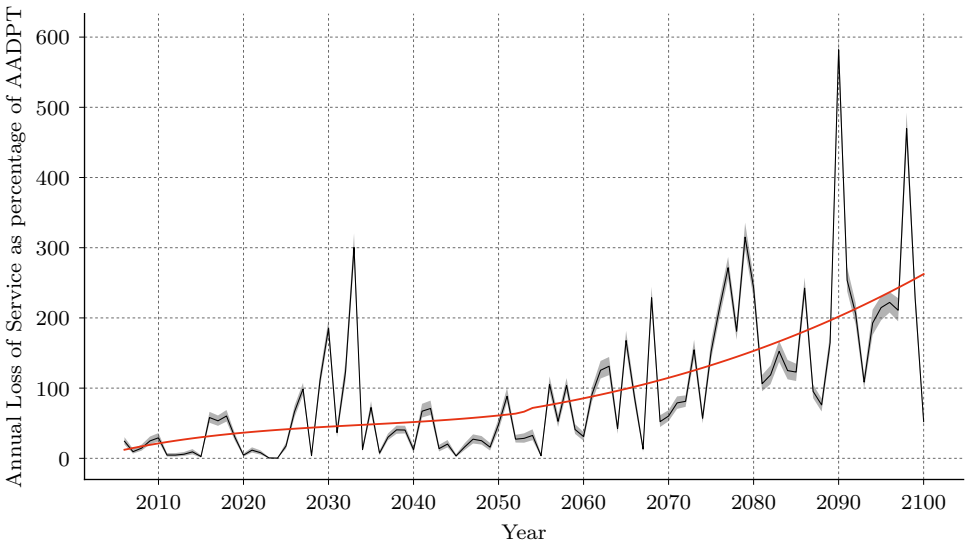


Figure 3.8: Time series of the estimated annual total LOS for the 2006-2100 period. The solid black line shows the expectations of the aggregated distributions. The shaded area shows the 5% to 95% range across the aggregated distribution. The solid red line is obtained by smoothing the solid black line with a Savitzky–Golay filter (Savitzky and Golay).

3.3.4 Shock-disruption Relationships

In this case study, the weather profiles are regarded as a form of external shock imposed on the railway system. The direct effect of such shock is asset failures and removal of edges from the asset layer. Service disruption occurs when the railway system fails to deliver a flow along its original path due to asset failures. Infrastructure systems are believed to be complex systems

that often exhibit non-linearity (Punzo et al., 2018). The railway system in this case study is no exception. The incidence of asset failure doesn't increase linearly with rising temperatures. As depicted in 3.9a, there exists a threshold temperature, beyond which a sudden surge in asset failures is observed. Additionally, the magnitude of both initial and cumulative service loss does not scale linearly with the number of asset failures, as shown in Figure 3.9d.

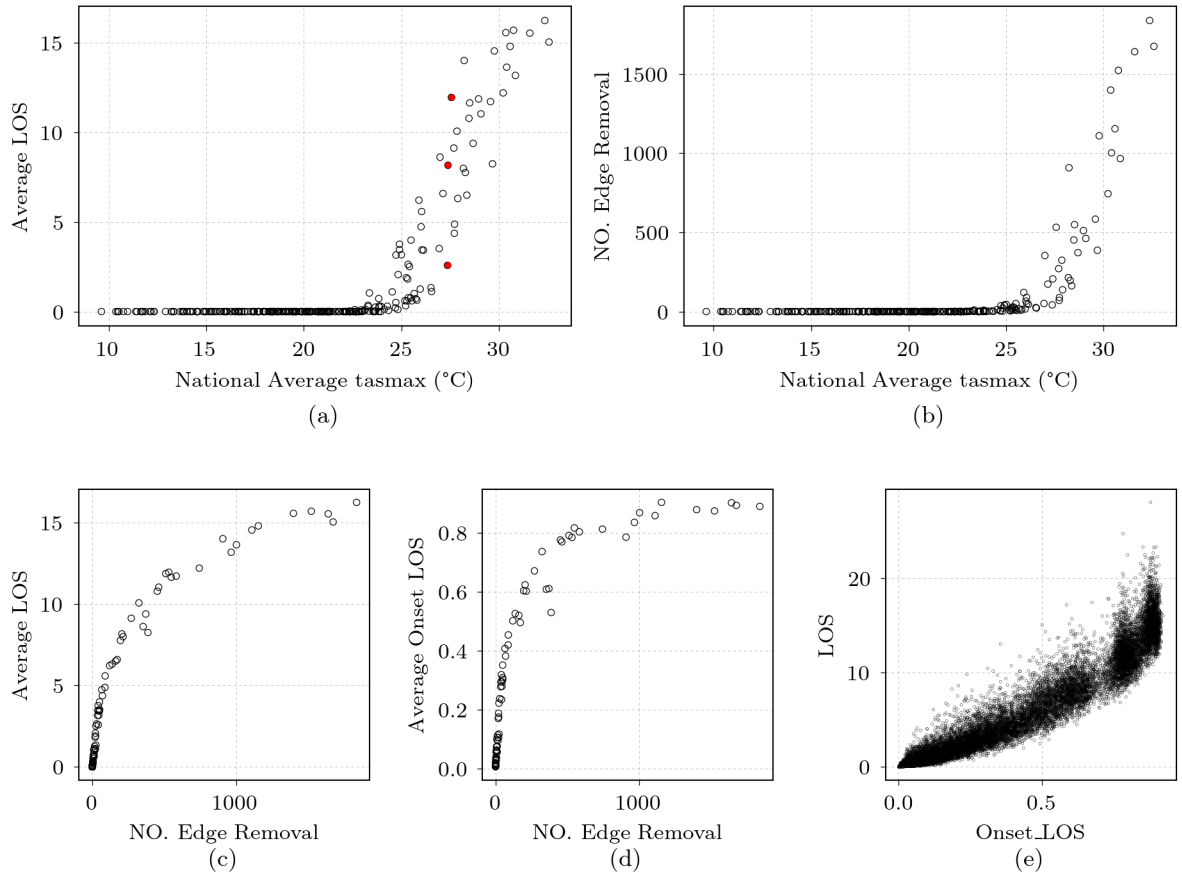
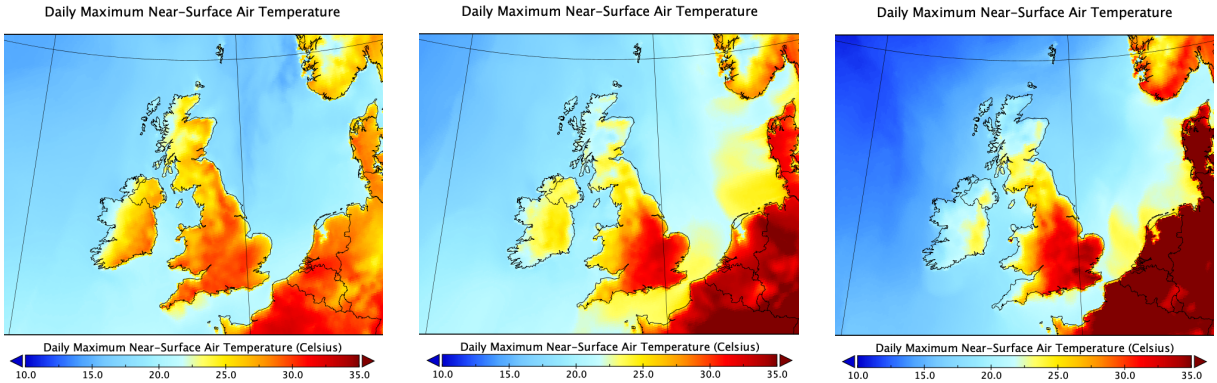


Figure 3.9: Relationship between (a) The average of measured cumulative loss of service and the national average tasmax; (b) the expectation of the number of edge removed Ψ and the national average tasmax; (c) The average of measured cumulative loss of service and expectation of the number of edge removal, Ψ; (d) the expectation of the number of edge removed Ψ and the average of loss of service at the onset of disruption; (e) the average of cumulative loss of service and the onset loss of service. The national average tasmax is the average of the tasmax across the asset layer.

Figure 3.9a shows this shock-disruption relationship, with the magnitude of the external shock measured by the national average tasmax and the severity of the disruption indicated by the simulated LOS. The plot uses the simulation results of the example days, whose expected number of failure, Ψ, is greater than 1. For each day, 250 failure scenarios were sampled and for each failure scenario, a LOS was calculated as described in Section 3.3.2. The average of the 250 simulated LOS is plotted against the national average tasmax of the corresponding example day. As the temperature increases, the increase of the consequential LOS presents a threshold behaviour with an apparent surge as the national average tasmax reaches 25 °C. Following the surge, there is a steady increase with the increasing temperature followed by a trend to plateau. This suggests that changes in the climate system can lead to some disproportional disruptions

to the infrastructure systems.

This shock-disruption relationship is then further broken down. The shock-disruption relationship in Figure 3.9a is in fact a combined effect of the shock-damage relationship in Figure 3.9b and the damage-disruption relationship in Figure 3.9c. Figure 3.9b shows the relationship between the magnitude of the shock and the extent of damage caused, where a surge happens at a 25 °C national average of tasmax followed by a linear increase with increasing temperature. Figure 3.9c shows the relationship between the extent of damage caused (number of edges removed) and the severity of service disruption (LOS). The plot has a sharp increase as few edges are removed and then keeps on growing at a slower rate. The damage-disruption relation in Figure 3.9c is then broken down into Figure 3.9d and Figure 3.9e. Figure 3.9d shows the relationship between the number of edges removed and the scale of disruption caused at the onset of the event. Figure 3.9e shows the relationship between the onset service disruption (without any rerouting or repairing) and the cumulative loss of service. The plots suggest that the non-linear relationship between the number of edge removal and the onset of service disruption is one of the main contributors to the overall observed system non-linearity.



(a)
15 August 2091
Avg tasmax = 27.4°C
Avg LOS = 2.6

(b)
11 August 2080
Avg tasmax = 27.4°C
Avg LOS = 8.2

(c)
30 June 2086
Avg tasmax = 27.6°C
Avg LOS = 12.0

Figure 3.10: Mapping of the tasmax for three days with similar national average tasmax but different LOS.

In Figure 3.9a, three data points are marked red. For those three example days, the tasmax averaged near 27.5°C across the whole railway track network. Figure 3.10 maps the tasmax for those three days. The left map corresponds to the bottom dot, the middle figure corresponds to the dot in the middle and the right figure corresponds to the dot at the top. With a similar average tasmax, the simulated LOS varies. From the map, it is firstly noticeable that the temperature distribution is evener in Figure 3.10 (a) than the other two. In Figure 3.10 (b) and (c), temperature at the south and east of England is marginally higher than the rest of the country, which is where a large amount of flows in the network are travelling pass as shown in Figure 3.4.

3.3.5 Climate-based, Random and Targeted Failures

As discussed in Chapter 2, neither the random failure scenarios, where failures happen randomly across the network nor the targeted attack strategy, where network component failures are targeted to simulate the most extensive possible disruption, may be able to capture the feature of weather-related disruptions. Therefore, to compare the effects of these against climate-based disruption, two separate sets of simulations, one implementing the random failure scenarios and the other implementing the targeted attack strategy, are carried out. For a given day, the expected number of edge failures, Ψ, is calculated using Equation 3.17. For the random strategy, in each Monte Carlo run, Ψ number of edges are randomly selected and removed from the network, followed by the same rerouting and repairing efforts with the LOS calculated using the same approach. The targeted attack strategies used in literature, to some extent, aims at capturing events like terrorist attacks or some theoretic worst case scenarios. Here, the simplest form of targeted attack is used, where the first Ψ edges with the most traffic are removed from the network, followed by the same resilience assessment procedure.

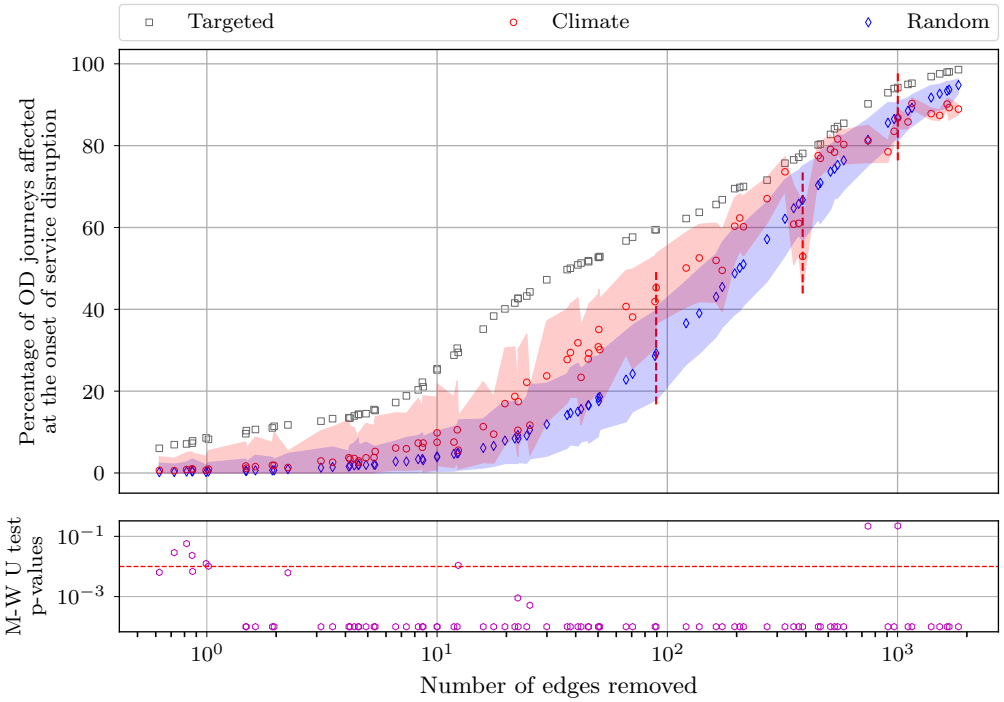


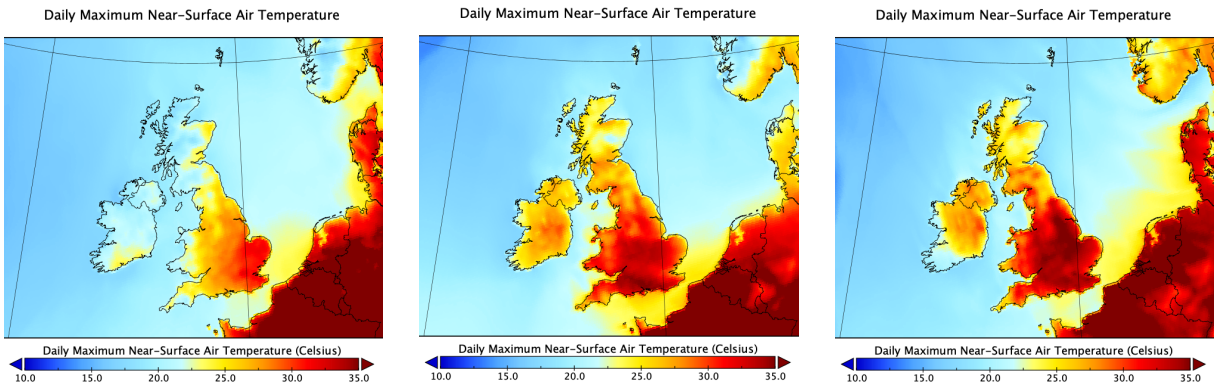
Figure 3.11: Comparing the the initial onset loss of service between the climate-based, random and targeted failure scenarios. Upper - the percentage of OD journey interrupted at the onset of disruption without any rerouting or repairing effort against the number of nodes removed. The sample size is 250. The shaded area shows the 2.5% to 97.5% range. Markers are the means of each sample. Lower - p-values of the Mann-Whitney U test between samples from climatic and random failure scenarios vs. the number of nodes removed. Any p-value smaller than 0.0001 is replaced with 0.0001. The red horizontal line is p=0.01, the confidence level adopted.

Figure 3.11 shows the percentages of OD flows interrupted at the onset of component failure without any rerouting or repair attempts for all three strategies. When the same number of edges

are removed from the asset layer network, the strategy that results in a higher percentage of OD disruption is believed to be more disruptive. The plot shows that climate-based strategy tends to sit between random and targeted. In particular, for the extent of the failed portion of the network edges intermediate between a sparse and a total number of failures, the climate-based strategy is more disruptive than random and less disruptive than targeted. For an extremely high number of edges involved, the disruption of a random strategy would exceed the one generated by the climate-based though there is only a slight difference between the three strategies.

The Mann-Whitney U test (Mann and Whitney, 1947) is carried out between samples of the climatic-based and random failure scenarios. This specific test was chosen because more than half (45 out of 84 for the climate-based scenarios, and 61 out of 84 for the random scenarios) of the sampled LOS fails the normality test with a significance level of 0.01. The null hypothesis for the Mann-Whitney U test is that the distributions of the two samples are identical and the alternative hypothesis is that they are not identical. With a significance level of 0.01, the null hypothesis is rejected in 76 out of the total 84 instances. The results of the statistical tests highlight how different the climate-based failure scenarios are from the random disruptions.

Three vertical dashed lines are marked in Figure 3.11. The leftmost dashed line marks the day with the most significant difference between the onset LOS simulated from random failure scenarios and climate-based failure scenarios. Figure 3.12(a) maps the tasmax for the day where high temperatures are observed on the east coast up to 35°C and the west coast north is around 15°C. The second dashed line denotes the day where climate-based failure scenarios seem less disruptive than random ones. Figure 3.12(b) shows the tasmax for that day, where the 25°C threshold (as observed in Figure 3.9 b) is breached across the whole network. The third dashed line represents a day with nearly zero difference between the onset LOS obtained with random and climate-based failure scenarios. This observation leads to the hypothesis that the uneven probabilities of failure, resulting from geographically varying temperatures, could be the reason for the differences between climate-based and random failure scenarios.



(a) 27 July 2088 (b) 07 August 2080 (c) 30 August 2068

Figure 3.12: Mapping of the tasmax for three days that the climate-based failure scenarios turn out to be (a) more disruptive, (b) less disruptive, and (c) as disruptive as the random failure scenarios.

Figure 3.13 shows the distribution of the measured cumulative loss of service against the number of edges removed for the three strategies. Though discernible differences exist between the LOS

sampled from the climate-based failure scenarios and the random strategy, the Mann-Whitney U tests are carried out between the samples. Again, this specific test was chosen because the majority (167 out of 168) of the sampled LOS failed the normality test with a significance level of 0.01. The null hypothesis for the Mann-Whitney U test is that the distributions of the two samples are identical and the alternative hypothesis is that they are not identical. With a significance level of 0.01, the null hypothesis is rejected in 74 out of the total 84 instances. It suggests that the LOS sampled from the climate-based failure scenarios are statistically different from those sampled with random failure in most cases. In contrast, the targeted strategy sits distinctly above both the random and the climate-based for all extents of edge failure. Overall, the system suffers a higher level of loss of service under the climate-based failure scenarios than the random one when the same number of edges are removed from the network. It means the system has a higher level of resilience toward random failure scenarios than climate-based.

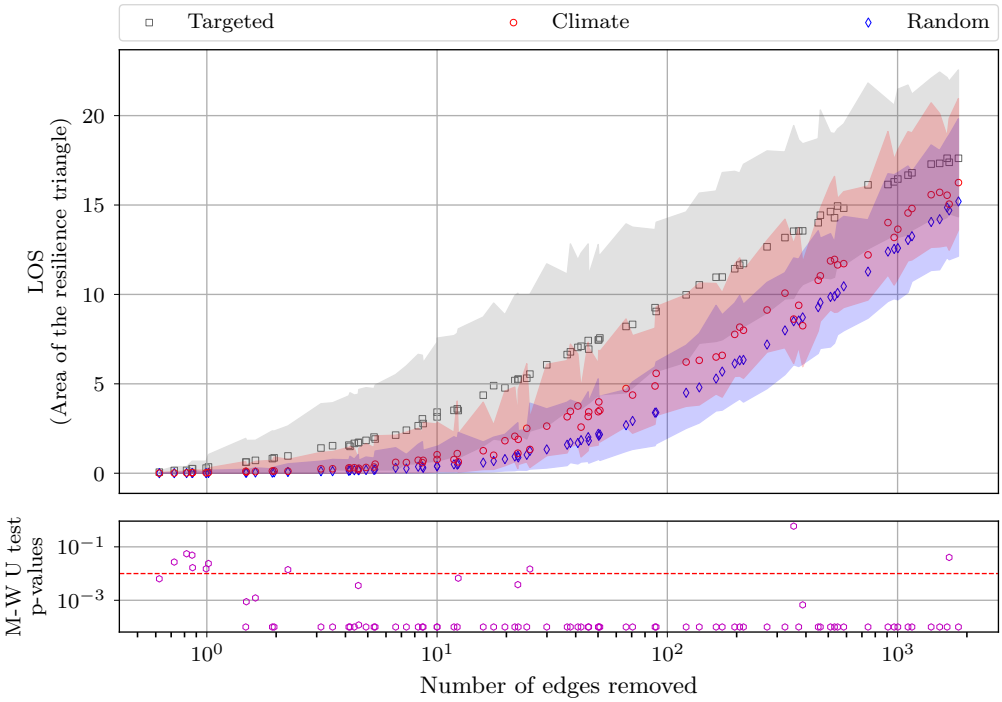


Figure 3.13: Comparing the cumulative loss of service between the climate-based, random and targeted failure scenarios. Upper - the measured LOS against the number of edges removed. The sample size is 250. The shaded area shows the 2.5% to 97.5% range. Markers are the means of each sample. Lower - p-values of the Mann-Whitney U test between samples from climate-based and random failure scenarios against the number of nodes removed. Any p-value smaller than 0.0001 is replaced with 0.0001 for visualization purposes. The red horizontal line is p=0.01, the confidence level adopted.

3.3.6 Case Study Discussion

As presented in Section 3.3.3, Section 3.3.4, and Section 3.3.5, the case study points to three key findings. 1) There is an overall increasing trend in the estimated annual total LOS over the 95-year period, where increased frequency and severity of extreme events drive the average LOS upwards.

2) The severity of the disruptions caused, measured with the cumulative LOS, increases non-linearly with the increasing magnitude of the external shock applied. This non-linearity is a combined effect of the threshold behaviour between the national average temperature and the number of edges removed (Figure 3.9b) and the nonlinear relationship between the number of edges removed and average onset disruption (Figure 3.9d). The reason behind the nonlinear relationship between the number of edges removed and average onset disruption is down to the nature of complex systems, where the behaviour or response of the system is distinct from the combined behaviour or response of its components (Punzo et al., 2018). The key message here is that climate change impact assessment should not stop at the assessment of asset-level physical damage but need to be taken to system-level functional loss.

3) Random failure models tend to overestimate the network's resilience compared to climate-based failure scenarios. When climate-based disruptions are considered, the network function degradation is more severe than in the random failure scenarios. This finding supports the working hypothesis that infrastructure systems may behave differently under different failure scenarios and evidences the need to use realistic weather profiles generated by climate models.

Further exploration of the reason behind this would require more simulations with various failure scenarios to test any proposed hypothesis and is beyond the scope of this work. The most remarkable difference between the random failure scenario and the climate-based failure scenario is the unevenness of the probabilities of failure. In climate-based failure scenarios, the probabilities of failure vary among assets due to differences in local weather. In random failure scenarios, when a specific number of network components are selected for removal, all network components have an equal probability of being chosen. The probability of removal is the same for all assets. Moreover, assets of close geographical proximity are likely to have similar probabilities of failure as the weather is more likely to be similar. Future works could look at if such spatial correlations of weather-induced failures leads to clusters of assets, within close geographic proximity, being affected simultaneously and if it is the cause of the differences between the climate-based failure scenarios and the random ones.

These findings from the case study are considered not specific to the infrastructure model and climate model used. Regardless of the specific choices for this case study, the method remains valid as it does not rely on specific features of the dataset chosen. As for the climate dataset, to completely prove the universality of the phenomenon observed, simulations of output from all climate models would be required and is of huge computational cost. The evidence we found about the difference between random and climate-induced failures holds across the vast majority of the days examined. The conditions found in our sample could be found in any dataset where only the distribution of the magnitude and the frequencies change.

The case study is carried out as to showcase how the proposed failure initiation method can be integrated into a resilience assessment framework for climate resilience assessment purpose. Several simplifications and assumptions are made in the case study in the absence of relevant information. Some of these simplifications and assumptions are inherently difficult to justify and validate. Yet it is the most general choice could be made and can be further improved in

future applications of the proposed method.

This case study only concerns track buckling in extreme heat while the industry's immediate response to high temperatures is to reduced train speed. Depending on the specification and conditions of tracks, temporary speed limits are implemented when the air temperature exceeds certain thresholds. Lower speed means lower forces on the tracks and hence lower probabilities of track buckling. For example, for a length of 'undisturbed, fully ballasted and consolidated' track, a watchman will be in place to monitor the tracks when air temperature exceeds 35°C, a 30/60 mph speed restriction will be applied when the air temperature exceeds 39°C, and a 20 mph speed restriction for 42°C. While for a length of track where 'ballast generally full between sleepers and on shoulders, but not consolidated', the corresponding critical air temperature is 24°C, 26°C, ad 27°C. The reduced speed means delays and then cancellations of fractions of schedule, which causes service disruptions. Further, heat does not just make the track buckle. It also affects the functioning of other rail assets such as the signalling electronics and the tensioning of the overhead lines. Service disruption happens even before the occurrence of any track bucking. Future case studies could introduce multiple asset layers to consider the effects of high temperature from the multiple aspects discussed above to better understand the impacts of global warming on the railway system.

The asset layer is an undirected network where there only exists one link between two nodes (stations). In reality, there may exists multiple railway lines between busy routes. The case study does not address the effects of multiple track lines. Multitrack models developed in literature could be used to improve the representativeness of the system model. If the effects of multiple tracks between lines are considered, when subjected to the same weather profile, in principle, there could be more asset failures compared to the current model, meaning the surge shown in Figure 3.9b) would be more sudden. For the relationships plotted in Figure 3.9c) and Figure 3.9d), the inclusion of the effect of multitracks means when the same number of edges are removed, the LOS and the onset LOS could be smaller than the current model. Though changes are possible, it is unlikely that the non-linear relationships observed would change to linear ones.

When assigning local weather parameters, only one value and hence one probability of failure is assigned to any given edge. Yet edges in the asset layer are of various length and longer edges are in fact of higher probability of experiencing a buckled track as they have more points of potential failure. This compromise can be improved by dividing edges into segments of appropriate length with each segment having its own failure probability. Introducing the variation of edge length into the formulation would only increase the variation, or unevenness, between the probabilities of failure for edges. As hypothesised above, increased unevenness between probabilities of failures could lead to higher LOS.

The rerouting algorithm is a modified Edmond-Karp maximum flow minimum cost algorithm. At each time step, the algorithm implements 5 rounds of path searching. At each round of path search, it searches for the shortest path that, 1) all edges forming the path has available spare capacity; and 2) is within twice the geographical length of the original path for each disrupted

OD pair. The probability of repair at the moment is kept at a 50% chance for all damaged assets without any consideration on amount of repairing resources available. The choice of '5 rounds', 'twice the length' and '50% chance' are all arbitrary values set in order to complete this prototype of the proposed assessment processes. Settings related to the rerouting and repairing algorithm only affects the speed of recovery but not the onset LOS. Therefore, the finding that the onset service disruption does not scale linearly with the increasing temperature remains valid.

Moreover, the long running time and makes it difficult to determine the number of realisations for the Monte Carlo simulation. The sample size of 250 is primarily determined by the 96h time limit of batch jobs submitted to the university's high-performance computing services. Each example day with a given failure strategy (climate-based, random, or targeted) is submitted as a batch job and all batch jobs for a given strategy is submitted together as an array job. With the current sample size of 250, the longest CPU time of a batch job is 89 hours and 20 minutes. 250 is the maximum number of samples the HPC would allow without risking job abortion. In this regard, it is checked that: 1) if N=250 is enough to carry out the statistical tests; 2) if N=250 is a good number for the Monte Carlo simulation.

For the statistical testing aspect, the current sample size of 250 is sufficient for the statistical tests that compare the climate-based and random failure scenarios. For 33 pairs of sampled onset LOS that are both approximately normal (significance level 0.001), a power analyses (alpha=0.05, power=0.8) was performed to determine the sample size needed. For 24 pairs of sampled onset LOS that are both approximately gamma distributed (Kolmogorov-Smirnov test for goodness of fit, significance level 0.001), the required sample sizes are calculated according to Cundill and Alexander (2015) (alpha=0.05, power=0.8). The histogram of the calculated sample size is presented in Figure 3.14.

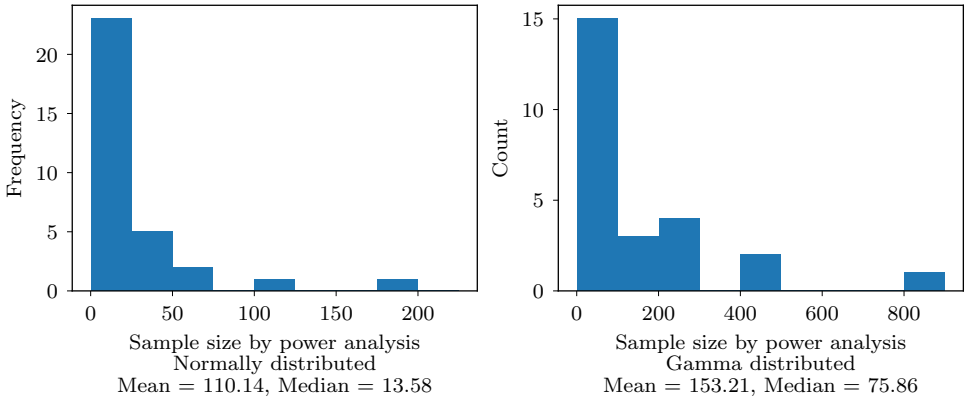


Figure 3.14: The histograms of calculated sample size with power analysis. Left - the calculated required sample size for the normally distributed pairs. Right - the calculated required sample size for the gamma distributed pairs.

For the Monte Carlo simulation, there is no incentive to sample more than 250. Extra 750 simulations were carried out on estimating only the onset disruptions, where only Q(t) at t=0 is calculated without any rerouting and repairing calculations. Figure 3.15 plots the difference

between the percentage of journey interrupted obtained from N=1000 and the percentage of journeys interrupted with N=250. The histogram shows that the differences for the climate-based group ranges from -0.88% to 0.51%. The difference for the random group ranges from -0.43% to 0.87%. Therefore, N=250 is sufficient to reveal the overall trends and the simulation results generated with N=250 is fully capable of supporting the findings.

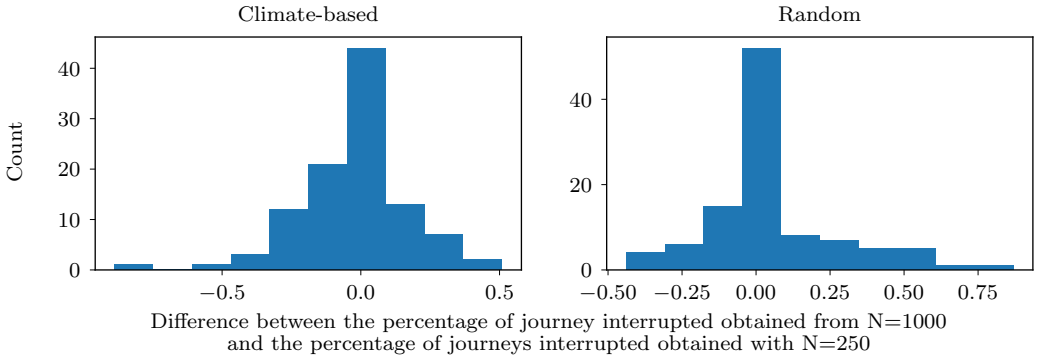


Figure 3.15: The difference between the percentage of journey interrupted obtained from N=1000 and the percentage of journeys interrupted with N=250, for the climate-based scenarios and the random removal scenarios.

3.4 Chapter Conclusion

This chapter presented a method to generate network failure scenarios and estimate system disruptions using climate change research data for the assessment of the infrastructure system's climate resilience. The case study attempts to quantify the resilience of Great Britain's railway passenger transport system to high-temperature-related track buckling under the RCP8.5 climate change scenario. Findings from the case study support the two arguments that motivate this framework's proposal: 1) random failure models tend to overestimate the network's resilience; and 2) the quality of services degrades non-linearly with the magnitude of the asset failures and the external weather parameter. Together, they prove the need for linking climate change resilience assessment to system-level functional loss as opposed to a count of the failed nodes or the identification of a threshold for network fragmentation. With the framework in this chapter able to quantify the climate-related service disruption, next chapter considers how to minimise this service disruption.

Chapter 4

Asset upgrade strategy towards climate resilience

4.1 Introduction

This chapter aims to address the second research question and considers the problem of finding the optimal or best possible resource allocation between infrastructure assets to minimise the system's service disruptions under plausible future extreme weather events with a finite budget.

Section 4.2 formulates the resource allocation problem as a constrained optimisation problem using the same railway network model with the heat-related hazard as in the previous chapter. The objective function is set to maximise the amount of service retained (Equation 4.17), or alternatively to minimise the onset loss of service (Equation 4.22), for a given weather profile. Both versions of the objective functions are formulated as a function of the key input variables, namely the local weather parameter, the design load, the variance in the design load, a link-route incidence matrix, and the daily OD pair-wise demand.

Section 4.3 presents a network metric-based resource allocation strategy and the development of three candidate network metrics based on the edge betweenness centrality (Equation 4.28, Equation 4.32, and Equation 4.36). Using a simple test network model, Section 4.4 uses the Lagrange multiplier method to solve the optimisation problem analytically. After the above implementations of the two distinct approach, Section 4.5 presents a rather interesting observation, where Equation 4.119 and Equation 4.121 is found identical. This finding shows the consistency between the network metric developed in Section 4.3 and the solution from Section 4.4. Figure 4.1 shows the structure of this chapter.

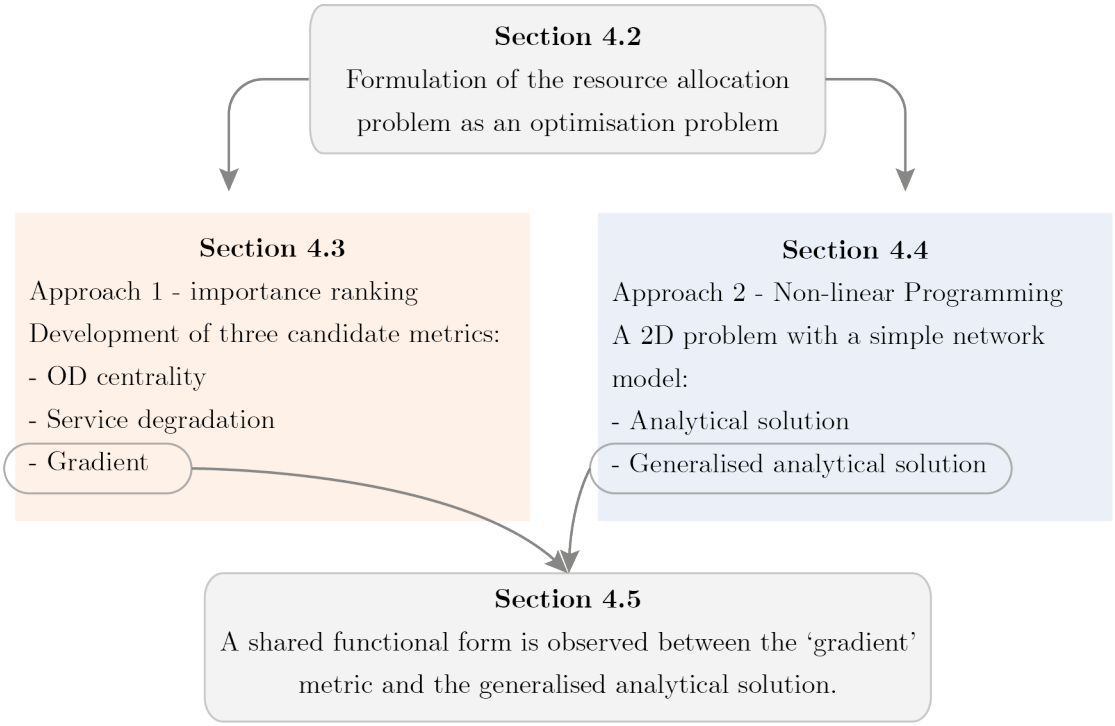


Figure 4.1: The structure of Chapter 4

4.2 Formulation of the Resource Allocation Problem

4.2.1 The System Model

Consider the same network model as in Chapter 3, with an asset layer, $\mathcal{G}_\alpha = \{\mathcal{V}_\alpha, \mathcal{E}_\alpha\}$, and an OD layer $\mathcal{G}_\phi = \{\mathcal{V}_\phi, \mathcal{E}_\phi\}$. The asset layer represents the network of the physical assets in the infrastructure system, such as stations and railway lines. The flow layer represents the service provided by the infrastructure system, such as transporting passengers and commodities.

The node set of the OD layer is a subset of the asset layer, $\mathcal{V}_\phi \in \mathcal{V}_\alpha$, including the origin and destination nodes and excluding some intermediate connection nodes. The edge sets of the two layers are two separate sets, $\mathcal{E}_\phi \notin \mathcal{E}_\alpha$. In this chapter, the edge sets are assumed to be a list where the edges are in a specific order so that vectors associated with the edges' properties, for example, the unit costs of upgrade edges and the amounts of service flow between the OD pairs, are of the same order.

Edges in the OD layer represent the flow of services. Those service flows are assumed to rely on the edges in the asset layer. Moving commodities by rail from the origin node to the destination node requires all track segments along its routes to be in working condition. Those track segments along its route form a list of assets needed to deliver this service. For example, in the simple network illustrated in Figure 4.2, OD flow F_1 relies on the functioning of edge e_1, e_2, and

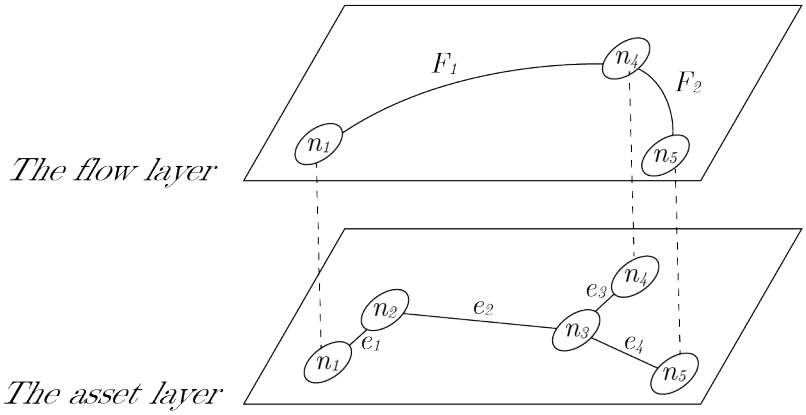


Figure 4.2: A simple network illustration for the bi-layer structure.

e_3 in the asset layer. OD flow F_2 relies on the functioning of edge e_3 and e_4 in the asset layer. Further, each OD pair is assumed to have one specific route in the asset layer in the network model here. Hence an *OD pair* refers to an edge in the flow layer, whilst an *OD path* refers to an ordered list of edges in the asset layer that the OD pair relies on to deliver the OD flow.

This dependent relationship between the asset layer and the flow layer is captured by a link-route incidence matrix H. The number of rows in H equals the number of edges in the asset layer and the number of columns equals the number of OD pairs in the flow layer. Therefore H is of shape $|\mathcal{E}_\alpha| \times |\mathcal{E}_\phi|$. $H_{i,j} = 1$ denotes that edge $e_i \in \mathcal{E}_\alpha$ belongs to the OD path of OD pair $OD_j \in \mathcal{E}_\phi$. Otherwise $H_{i,j} = 0$. For example, the link-route incidence matrix H for the simple network in Figure 4.2 is:

$$\begin{array}{c} \\ e_1 \\ e_2 \\ e_3 \\ e_4 \end{array} \begin{array}{c} \begin{array}{cc} F_1 & F_2 \end{array} \\ \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \end{array}.$$

The first column is $[1, 1, 1, 0]$, meaning OD flow F_1 relies on edge e_1, e_2, and e_3 in the asset layer but not edge e_4. The second column is $[0, 0, 1, 1]$, meaning OD flow F_2 relies on edge e_3 and e_4 but not e_1 and e_2.

4.2.2 The Objective Function

As discussed in Section 2.3.1, this thesis recognise that the ultimate purpose of infrastructure systems is to provide services and the resilience goal for infrastructure systems should be to minimise service disruption. As defined by the UK National Infrastructure Commission of the UK and the National Infrastructure Advisory Council of the USA, the resilience of infrastructure systems is associated with several distinct sub-aspects, including robustness - the ability to resist and absorb external shocks, and recoverability - the ability to restore services quickly (National

Infrastructure Commission, 2020; Berkeley et al., 2010). Referring to the concept of resilience and the resilience triangle discussed in Section 2.3.1, the area of the resilience triangle primarily depends on the scale of the onset disruption (base of the triangle, measured along the vertical axis) and time taken to restore full functionality (height of the triangle, measured along the horizontal axis).

As different from the work in Chapter 3, work in this chapter focuses on minimising the onset disruption, which ultimately will lead to a minimised cumulative service loss. Therefore, the objective function for the optimisation problem can be minimising the onset loss of service, $Q(t = 0)$. Alternatively the objective function can be maximising the amount of remaining service, $1 - Q(t = 0)$. The previous chapter uses the Monte Carlo simulation to estimate the onset LOS for a given weather profile. This chapter aims to complete the formulation and express the expectation of the amount of reaming services,or the expected flow, E, as a function of the local weather parameter ($\boldsymbol{\omega}$), the design load ($\boldsymbol{\mu}$), the variance in the design load ($\boldsymbol{\sigma}$), the link-route incidence matrix (H), and the daily OD pair-wise demand (F). To begin with, for a given weather profile, the local weather parameters for all edges:

$$\boldsymbol{\omega} = \{\omega_i, \cdots\},\ i = 1, 2, \cdots |\mathcal{E}_\alpha|. \tag{4.1}$$

A fragility function is used to convert the weather profile to probabilities of asset failure. This follows the approach for failure generation as in the previous chapter and uses the same functional form for the fragility function[1], as:

$$p_i = \frac{1}{2}\left[1 + \text{erf}\left(\frac{\omega_i - \mu_i}{\sigma_i\sqrt{2}}\right)\right], \tag{4.2}$$

where $p_i \in [0, 1]$ is the probability of failure for edge e_i, which could take any value between 0 and 1; ω_i is the weather parameter (e.g. temperature, wind speed, and precipitation) considered and can take different climate variables depending on the climate hazard considered; and μ_i and σ_i are the shape-control parameters for the fragility function for asset e_i and are also associated with the asset's condition.

The probabilities of individual asset failure can be mapped to probabilities of OD flow interruption. An OD path is assumed to only function when all edges in its OD path function. Therefore, the probability of success for an OD path OD_j equals to the probability of all of its dependent edges succeeding. For example, in the simple network, the probability of success for OD flow F_1 is $(1 - p_1)(1 - p_2)(1 - p_3)$ and the probability of success for OD flow F_2 is $(1 - p_3)(1 - p_4)$, where p_i is the probability of failure for edge e_i and $1 - p_i$ denotes the probability of success for a single edge calculated as one minors its probability of failure.

The link-route incidence matrix is used to tidy up the above two expression. The link-route incidence matrix is defined as $H_{i,j} = 1$ denoting that edge $E_i \in \mathcal{E}_\alpha$ belongs to the OD path of

[1] erf denotes the error function $\text{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$

OD pair, $OD_j \in \mathcal{E}_\phi$. Otherwise $H_{i,j} = 0$. The expression:

$$\prod_i (1 - p_i) \tag{4.3}$$

denotes the product[2] of $(1 - p_i)$ for all i. Ideally, for a given OD flow F_j, the above expression should only include edges that F_j depends on but excludes the failure probabilities of irrelevant edges. To achieve this, p_i is multiplied with $H_{i,j}$. Through this multiplication, if edge e_i is a dependent edge for OD flow F_j, when calculating the probability of success for F_j, the term $1 - H_{i,j}p_i$ would be $1 - 1 \cdot p_i = 1 - p_i$. Otherwise the term would be $1 - 0 \cdot p_i = 1$. Again, use the simple network as an example. The probability of success for OD flow F_1 is

$$(1 - 1 \cdot p_1)(1 - 1 \cdot p_2)(1 - 1 \cdot p_3)(1 - 0 \cdot p_4) = (1 - p_1)(1 - p_2)(1 - p_3), \tag{4.4}$$

and the probability of success for OD flow F_2 is

$$(1 - 0 \cdot p_1)(1 - 0 \cdot p_2)(1 - 1 \cdot p_3)(1 - 1 \cdot p_4) = (1 - p_3)(1 - p_4). \tag{4.5}$$

Therefore, the probability of success for OD path F_j is the product of $(1 - H_{i,j}p_i)$ for all i:

$$\prod_i (1 - H_{i,j}p_i). \tag{4.6}$$

With F_j denotes the amount of demand along OD path OD_j, the expected delivery for this OD pair, E_j can be obtained by multiplying F_j with its path's probability of success:

$$E_j = F_j \prod_i (1 - H_{i,j}p_i). \tag{4.7}$$

With the probabilities of failure for each individual edges multiplied together in the above functional expression, it considers the interactions between the probabilities of failures for different edges. The effectiveness of reducing the probabilities of failure for one edge is connected to the probabilities of failure of the edges on the OD path F_j. The expected delivery across the whole system is therefore sum of the expected delivery for all of the OD pairs:

$$\begin{aligned} E &= \sum_j E_j \\ &= \sum_j F_j \prod_i (1 - H_{i,j}p_i) \\ &= \sum_j F_j \prod_i \left(1 - H_{i,j}\frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{\omega_i - \mu_i}{\sigma_i\sqrt{2}}\right)\right)\right). \end{aligned} \tag{4.8}$$

[2] $\prod$ is the notation for the product of multiple items.

4.2.3 The Resource Allocation Problem

The optimisation problem in this chapter is to find the optimal or best possible resource allocation between infrastructure assets to minimise the system's service disruptions under plausible future extreme weather events with a finite budget. As Equation 4.8, the section above formulates the expected delivery of for a given weather profile as a function of the local weather parameter ($\boldsymbol{\omega}$), the design load ($\boldsymbol{\mu}$), the variance in the design load ($\boldsymbol{\sigma}$), the link-route incidence matrix (H), and the daily OD pair-wise demand (F). The formulated expression will be used as the objective function for the optimisation model.

The optimisation model assumes that edges in the asset layer can be upgraded to a higher design load μ_i^* from its original design load μ_i by $\Delta\mu_i$:

$$\mu_i^* = \mu_i + \Delta\mu_i, \tag{4.9}$$

so that their failure probability will be reduced when exposed to the same weather parameter:

$$p_i* = \frac{1}{2}\left(1 + \text{erf}\left(\frac{\omega_i - \mu_i^*}{\sigma_i\sqrt{2}}\right)\right). \tag{4.10}$$

The vector $\boldsymbol{\Delta\mu}$ contains the amount of upgrade $\Delta\mu_i$ for all of the asset:

$$\boldsymbol{\Delta\mu} = \{\Delta\mu_i, \cdots\},\ i = 1, 2, \cdots |\mathcal{E}_\alpha|. \tag{4.11}$$

Upon such an upgrade, when the system is under the same external weather parameters $\boldsymbol{\omega}$, the expected delivery should now increased from E to E^*:

$$E^* = \sum_j F_j \prod_i \left(1 - H_{i,j}\frac{1}{2}\left(1 + \text{erf}\left(\frac{\omega_i - (\mu_i + \Delta\mu_i)}{\sigma_i\sqrt{2}}\right)\right)\right). \tag{4.12}$$

Such an upgrade would come with a cost. Let c_i denote the unit cost to increase the design load of edge e_i. The values of c_i could differ from edge to edge depending on asset conditions. The cost of upgrade edge e_i by $\Delta\mu_i$ is simply the product of the unit cost and the amount of increase in the design load, $c_i\Delta\mu_i$. The total cost of upgrading the whole network by $\boldsymbol{\Delta\mu}$ is therefore the sum of the cost for upgrading all of the edges:

$$\sum_{i=1} c_i\Delta\mu_i. \tag{4.13}$$

There is always a finite budget for infrastructure asset upgrades. Suppose a given budget is made available to upgrade the infrastructure system, which can be used for any assets. Let C denote this budget. The first boundary condition is that the total cost of the upgrade should be less than or equal to the given budget:

$$\sum_{i=1} c_i\Delta\mu_i \leq C. \tag{4.14}$$

The increase in design load should also respect some physical and practical limits. For example, the probability of track buckling is related to the amount of lateral resistance provided by ballasts and sleepers (Kish et al., 2013). The strength of the material that the sleepers are made of can only be improved to a certain extent. Also, if there is no limit on $\Delta\mu_i$, the expected delivery theoretically can be infinitely close to the demand, where every asset is over-designed and returns a near-zero probability of failure for any weather parameter. Taking other economic factors into account, there shall be an upper limit to which the design load can not exceed, meaning $\Delta\mu_i$ should be capped. The upper boundaries for asset upgrade depend on various factors such as the types of materials the physical assets are constructed of, the development and innovations in relevant fields, and economic considerations. Let $\Delta\mu_i^{ub}$ detonate the upper limit for the increment in design load for edge e_i. The second set of boundary conditions is:

$$\Delta\mu_i \leq \Delta\mu_i^{ub}, \quad \text{for } i = 0 \text{ to } i = |\mathcal{E}_\alpha|. \tag{4.15}$$

Finally, and most importantly, it is assumed that design loads can only be increased, even though from the point of circular economy, the idea of knocking down some existing (not-in-use) assets to enhance others is sound. However, determining the utilisation ratio for such conversion can be complicated. Further, $\Delta\mu_i$ is assumed to be a continuous variable that can take any value between the upper and lower boundaries, one can not partially knock down an asset and invest the partially taken-away design load into a different asset. Therefore, this work assumes that the design load can only be increased. The third boundary condition is:

$$\Delta\mu_i \geq 0, i = 1, 2, \cdots |\mathcal{E}_\alpha|. \tag{4.16}$$

With the objective function formulated as the expected delivery in Equation 4.12 and the three boundary conditions defined, the optimisation problem states: for a given budget C, find the vector $\boldsymbol{\Delta\mu}$ that maximise the expected delivery E^*.

$$\max_{\boldsymbol{\Delta\mu}} \sum_j F_j \prod_i \left(1 - H_{i,j} \frac{1}{2}\left(1 + \text{erf}\left(\frac{\omega_i - (\mu_i + \Delta\mu_i)}{\sigma_i\sqrt{2}}\right)\right)\right) \tag{4.17}$$

subject to:

$$\sum_{i=1} c_i \Delta\mu_i = C, \tag{4.18}$$

$$\Delta\mu_i \geq 0, \qquad \text{for all } i, \tag{4.19}$$

$$\Delta\mu_i \leq \Delta\mu_i^{ub}, \quad \text{for all } i. \tag{4.20}$$

In optimisation, sometimes an alternative objective function,which does not change the optimal solution, makes the problem easier to solve. For example, instead of minimising $f(x)$, an alternative objective function is to maximise $A - f(x)$, where A is a constant. The sum of all OD flows $\sum_j F_j$ is the amount of flows delivered in the network without any asset failures and is a constant for a given optimisation problem of a system model. Here, the objective function can

be replaced as minimising the service disruption, which is calculated as sum of all OD flows F_j minus the expected delivery E^*.

$$\sum_j F_j - E^* = \sum_j F_j \left(1 - \prod_i \left(1 - H_{i,j} \frac{1}{2} \left(1 + \text{erf}\left(\frac{\omega_i - (\mu_i + \Delta\mu_i)}{\sigma_i \sqrt{2}}\right)\right)\right)\right) \tag{4.21}$$

With the alternative objective function, the optimisation problem states:

$$\min_{\boldsymbol{\Delta\mu}} \sum_j F_j \left(1 - \prod_i \left(1 - H_{i,j} \frac{1}{2} \left(1 + \text{erf}\left(\frac{\omega_i - (\mu_i + \Delta\mu_i)}{\sigma_i \sqrt{2}}\right)\right)\right)\right), \tag{4.22}$$

subject to Equation 4.18 - 4.20

The two optimisation problems stated in Equation 4.17 and Equation 4.22 have the same optimal solution. Depending on the solution-finding algorithms used, one may be easier to solve than the other. With the optimisation problem formulated, the next two sections present two solution finding methods - one utilises the importance ranking of network components (Section 4.3) while the other relies on classical linear programming with the Lagrange multipliers (Section 4.4). The unification of the two methods is presented in Section 4.5.

4.3 Development of Ranking Strategies

Network resource allocation problems can often be seen as ranking problems. The underlying assumption is that components of high topological or functional importance, e.g. a vertex with many edges, or an edge that is part of the shortest path of many node pairs, should be prioritised. Network components of relatively high importance enjoy proportionally large amounts of resources. This type of network-metric-based resource allocation approach is often computationally tractable. The idea of the metric-based resource allocation is that the amount of upgrade for an edge is proportional to its calculated ranking metric:

$$\Delta\mu_i \propto \gamma_i, \tag{4.23}$$

where γ_i is the calculated metric for edge e_i.

Let χ be a constant such that the amount of upgrade for edge e_i can be calculated by simply multiplying χ with the calculated metric for edge e_i:

$$\Delta\mu_i = \chi\gamma_i. \tag{4.24}$$

To satisfy the first boundary condition, the total cost of upgrade for all edges need to equal to the given budget, C.

$$\sum c_i \Delta\mu_i = \sum c_i \chi \gamma_i = C \tag{4.25}$$

Now the value of χ can be found by rearranging Equation 4.25:

$$\chi = \frac{C}{\sum c_i \gamma_i}. \tag{4.26}$$

Put the found expression of χ back into Equation 4.24, the resource allocation solution can be found by calculating each $\Delta\mu_i$ as:

$$\Delta\mu_i = \gamma_i \frac{C}{\sum_{i=1} c_i \gamma_i} \tag{4.27}$$

The solutions generated this way naturally satisfy the boundary conditions $\sum c_i \Delta\mu_i = C$ and $\Delta\mu_i \geq 0$. The upper boundary constraint can be handled by resetting $\Delta\mu_i$ to $\Delta\mu_i^{ub}$ if $\Delta\mu_i > \Delta\mu_i^{ub}$. This resetting leaves $c_i * (\Delta\mu_i - \Delta\mu_i^{ub})$ amount of resources unused, which can be allocated by running the metric-based resource allocation again. More details of handling the upper boundary condition handling are presented in Chapter 5 Section 5.3.5.

However, the metrics developed and solutions generated is of various degrees of quality as pointed out by the literature review in Section 2.4.1. This section tries to derive new metrics from the objective function formulated in the previous section and aims to include the key input variables, namely $\boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma}, H$, and F in the formulated network metric. The betweenness centrality of a node/edge directly indicates how many origin-destination pairs would have their original shortest path disrupted if the node/edge fails. The edge betweenness centrality is the starting point of the development of a new network metric that factors in both the topological feature of the asset layer, the supply-demand feature of the service layer, and the climatic failure probability.

The following three subsections present three candidate network metric. These metrics are candidate metrics to be confirmed by the analytical solution in Section 4.4.

4.3.1 OD Centrality

The system model here is a multi-layer graph comprised of two layers - the asset layer, which captures the topology of the physical assets and the OD layer, which captures the pairwise supply-demand relationships. The OD pairs in teh flow layer can be seen as a subset of all possible pairwise relationships. The original edge betweenness centrality counts all pairwise relationships. However, in a practical context, not all pairwise relationships matter. To consider this issue, an importance measure, *OD-centrality*, is proposed, adapted from the classical edge betweenness centrality. This OD-centrality, $ODC(i)$, of edge $e_i \in \mathcal{E}_\alpha$, is measured as the sum of all of OD flows whose path passes edge e_i. Utilising the link-route incidence matrix, the OD-centrality for edge e_i is:

$$ODC(i) = \sum_j H_{i,j} F_j. \tag{4.28}$$

4.3.2 Service Degradation

The OD-centrality ranks edges based on the daily service demanded, which measures the components' importance in the system's fully functioning state without factoring in the probability of component failure. The OD-centrality metric has H and F in its expression, but does not have $\boldsymbol{\omega}, \boldsymbol{\mu}$, and $\boldsymbol{\sigma}$, which are the variables related to the probabilities of asst failure. One of the options to integrate the probability of edge failure into the formulation is to multiply the edge's OD betweenness centrality by its probability of failure. In this way, edges with high structural importance (high betweenness centrality) and a high probability of failure would have higher priority.

When there is zero asset failure, the expected amount of flow traverses edge e_i is:

$$ODC(i) = \sum_j H_{i,j} F_j, \tag{4.29}$$

which, essentially, is the OD betweenness centrality. For a given external weather parameter ω_i, the expected flow along OD path OD_j is the product of F_j multiplied by the probabilities of success of all of its dependent edges:

$$E_j = F_j \prod_i (1 - H_{i,j} p_i). \tag{4.30}$$

The amount of expected drop in the through flow regarding a given OD path OD_j is therefore F_j multiplied by one minus the product of the probabilities of success of all of its dependent edges:

$$F_j - E_j = F_j (1 - \prod_i (1 - H_{i,j} p_i)). \tag{4.31}$$

The amount of drop in flow throughput regarding edge E_i resulted from all OD paths that depend on edge e_i therefore is:

$$\begin{aligned} SD(i) &= \sum_j H_{i,j} (F_j - E_j) \\ &= \sum_j H_{i,j} F_j (1 - \prod_i (1 - H_{i,j} p_i)). \end{aligned} \tag{4.32}$$

4.3.3 Gradient

The last metric includes the probability of failure from a different angle. *'Gradient'* plays an essential role in non-linear programming problems. This metric is defined by the local gradient of the objective function, which is the cost/benefit of investing in an edge. For a given external weather profile, the expected flow along the OD path of an OD pair is:

$$E_j = F_j \prod_i (1 - H_{i,j} p_i). \tag{4.33}$$

In $\prod_i(1-H_{i,j}p_i)$, the term $(1-H_{i,j}p_i)$ is related to edge e_i. Separating the term from Equation 4.33 gives:

$$E_j = (1-H_{i,j}p_i)F_j \prod_{k,k\neq i}(1-H_{k,j}p_k). \tag{4.34}$$

The separation done in Equation 4.34 can be carried out to all expected OD flows E_j. With the separation in place, the total expected delivery across the network E is the sum of F_j for all j:

$$\begin{aligned} E &= \sum_j E_j \\ &= \sum_j (1-H_{i,j}p_i)F_j \prod_{k,k\neq i}(1-H_{k,j}p_k) \\ &= \sum_j F_j \prod_{k,k\neq i}(1-H_{k,j}p_k) - \sum_j H_{i,j}p_i F_j \prod_{k,k\neq i}(1-H_{k,j}p_k) \\ &= \sum_j F_j \prod_{k,k\neq i}(1-H_{k,j}p_k) - p_i \sum_j H_{i,j}F_j \prod_{k,k\neq i}(1-H_{k,j}p_k). \end{aligned} \tag{4.35}$$

Differentiating with respect to μ_i gives the 'gradient' of investing in edge e_i:

$$\frac{\partial E}{\partial \mu_i} = 0 - \frac{\partial p_i}{\partial \mu_i}\sum_j H_{i,j}F_j \prod_{k,k\neq i}(1-H_{k,j}p_k), \tag{4.36}$$

$$\frac{\partial E}{\partial \mu_i} = -\frac{\partial p_i}{\partial \mu_i}\sum_j H_{i,j}F_j \prod_{k,k\neq i}(1-H_{k,j}p_k). \tag{4.37}$$

By apply Equation 4.36 to all edges, the local gradients for all edges can be found:

$$\left\{\frac{\partial E}{\partial \mu_1}, \frac{\partial E}{\partial \mu_2}, \cdots, \frac{\partial E}{\partial \mu_i}\right\}. \tag{4.38}$$

The gradients calculated with Equation 4.36 is the last candidate metric considered in this work. The three metrics presented, the OD centrality metric (Equation 4.28), the service degradation metric (Equation 4.32), and the gradient metric (Equation 4.36) are the best guesses made. Next section presents the second method used to solve formulated constrained optimisation problems.

4.4 The Analytical Solution of A Test Model

This section attempts to directly solve the formulated optimisation problem with a simple network model. As stated in Equation 4.17 and Equation 4.22, the original problem is multidimensional, where the number of dimensions equals to the number of edges in the asset layer. It is impractical to directly solve the original optimisation problem analytically, considering real-world infrastructure systems typically have a large number of assets. Therefore a test model is used (Figure 4.3) in an attempt to obtain an analytical solution that expresses the optimal solution as a function of the key input variables - the local weather parameters (ω), the assets condition (μ), the unit cost (c), the pairwise service demands (F) and the link-route incidence

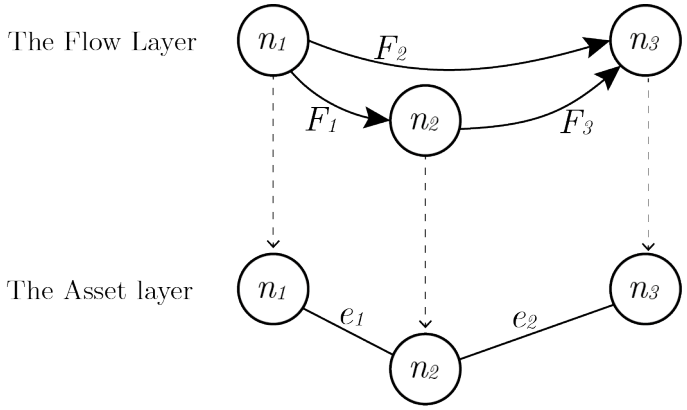


Figure 4.3: The network structure of the test model. The flow layer has 3 OD pairs. The asset layer is a network of 3 nodes and 3 edges.

matrix (H).

The asset layer (Figure 4.3, bottom) has 3 nodes and 2 edges. The service layer comprises three OD pairs (Figure 4.3, top). The current design loads are denoted as μ_1 and μ_2 for edge e_1 and e_2. The fragility function for each edge is:

$$p_1(\omega) = \frac{1}{2}\left(1 + \text{erf}\left(\frac{\omega - \mu_1}{\sigma_1\sqrt{2}}\right)\right), \tag{4.39}$$

$$p_2(\omega) = \frac{1}{2}\left(1 + \text{erf}\left(\frac{\omega - \mu_2}{\sigma_2\sqrt{2}}\right)\right). \tag{4.40}$$

The flow of services between OD pairs in the service layer is denoted as F_1, F_2, and F_3. The delivery of F_1 from n_1 to n_2 depends solely on the functioning of e_1, and therefore the first column of the route-link incidence matrix, H, is $[1, 0]$. The delivery of F_2 from n_1 to n_3 depends on the functioning of e_1 and e_2 and therefore the second column of H is $[1, 1]$. The third column is therefore $[0, 1]$. The route-link incidence matrix, H, is therefore:

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

4.4.1 The Objective Function

In this simple test model, $j \in \{1, 2, 3\}$ as there are three OD flows and $i \in \{1, 2\}$ as there are 2 edges in the asset layer. The total expected delivery can calculated as the sum of the OD flows

F_1, F_2, and F_3 multiplied by their corresponding OD path's probability of succeeding:

$$E = \sum_j F_j \prod_i (1 - H_{i,j} p_i) \tag{4.41}$$

$$\begin{aligned} =& F_1(1 - H_{1,1}\ p_1)(1 - H_{2,1}\ p_2) \\ &+ F_2(1 - H_{1,2}\ p_1)(1 - H_{2,2}\ p_2) \\ &+ F_3(1 - H_{1,3}\ p_1)(1 - H_{2,3}\ p_2), \end{aligned} \tag{4.42}$$

where,

$$p_1 = \frac{1}{2}\left(1 + \text{erf}\left(\frac{\omega - \mu_1}{\sigma_1\sqrt{2}}\right)\right), \tag{4.43}$$

$$p_2 = \frac{1}{2}\left(1 + \text{erf}\left(\frac{\omega - \mu_2}{\sigma_2\sqrt{2}}\right)\right). \tag{4.44}$$

Take the values of $H_{i,j}$ into the expression:

$$\begin{aligned} E &= F_1(1 - p_1) + F_2(1 - p_1)(1 - p_2) + F_3(1 - p_2) \\ &= (F_1 + F_2 + F_3) - p_1(F_1 + F_2) - p_2(F_2 + F_3) + p_1 p_2 F_2. \end{aligned} \tag{4.45}$$

Assuming the edges in the asset layer can be upgraded to higher design loads $\boldsymbol{\Delta\mu} = \{\Delta\mu_1, \Delta\mu_2\}$, with the upgrade, the reduced probabilities of failure, p_i^*:

$$p_1^* = \frac{1}{2}\left(1 + \text{erf}\left(\frac{\omega_1 - (\mu_1 + \Delta\mu_1)}{\sigma_1\sqrt{2}}\right)\right), \tag{4.46}$$

and

$$p_2^* = \frac{1}{2}\left(1 + \text{erf}\left(\frac{\omega_2 - (\mu_2 + \Delta\mu_2)}{\sigma_2\sqrt{2}}\right)\right). \tag{4.47}$$

The expected delivery with the upgrade, E^*, is:

$$E^* = (F_1 + F_2 + F_3) - p_1^*(F_1 + F_2) - p_2^*(F_2 + F_3) + p_1^* p_2^* F_2. \tag{4.48}$$

The expected initial loss of service (iLOS) at the onset of the disruption without any re-routing or repairing efforts, is:

$$\begin{aligned} iLOS &= (F_1 + F_2 + F_3) - E^* \\ &= p_1^*(F_1 + F_2) + p_2^*(F_2 + F_3) - p_1^* p_2^* F_2. \end{aligned} \tag{4.49}$$

In this case, the objective function is to maximise the expected delivery E^*

$$\max_{\boldsymbol{\Delta\mu}} E^* = (F_1 + F_2 + F_3) - p_1^*(F_1 + F_2) - p_2^*(F_2 + F_3) + p_1^* p_2^* F_2, \tag{4.50}$$

or minimise the initial loss of services (iLOS)

$$\min_{\boldsymbol{\Delta\mu}} iLOS = p_1^*(F_1 + F_2) + p_2^*(F_2 + F_3) - p_1^* p_2^* F_2, \tag{4.51}$$

subject to:

$$c_1\Delta\mu_1 + c_2\Delta\mu_2 = C, \tag{4.52}$$
$$\Delta\mu_1 \leq \Delta\mu_1^{ub}, \tag{4.53}$$
$$\Delta\mu_2 \leq \Delta\mu_2^{ub}, \tag{4.54}$$
$$\Delta\mu_1 \geq 0, \tag{4.55}$$
$$\Delta\mu_2 \geq 0. \tag{4.56}$$

Now a 2D version of the original optimisation problem problem is formulated, with a maximisation and a minimisation options for the objective function.

4.4.2 Direct Application of the Lagrange Multiplier Method

The Lagrangian multiplier method is one of the essential methods of finding the optimal solution for non-linear programming problem and is the theoretical foundation of various multiplier methods (Bertsekas, 2014). It is utilised here to obtain an analytical solution of the formulated 2D optimisation problem. Consider a basic optimisation problem:

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{subject to} \quad & g(x) = 0. \end{aligned}$$

The Lagrangian multiplier method first construct a function that combines the constraints and the objective function together:

$$L(x, \lambda) = f(x) + \lambda g(x), \tag{4.57}$$

which is called the Lagrangian function. To find the local minima, it need to find the x that makes the partial derivates of the Lagrangian function zero:

$$\frac{\partial L}{\partial x} = 0 \quad \text{and} \quad \frac{\partial L}{\partial \lambda} = 0. \tag{4.58}$$

The local minimums can be found by solve the above two equations together. For the simple resource allocation problem formulated in the previous subsection (Equation 4.51), the Lagrangian

function is defined as:

$$\begin{aligned} L(\Delta\mu_1, \Delta\mu_2, \lambda, \theta_1, \theta_2, \varphi_1, \varphi_2) =& p_1^*(F_1 + F_2) + p_2^*(F_2 + F_3) - p_1^* p_2^* F_2 \\ & - \lambda(c_1\Delta\mu_1 + c_2\Delta\mu_2 - m(c_1 + c_2)) \\ & - \theta_1(\Delta\mu_1 - \Delta\mu_1^{ub} + t_1^2) \\ & - \theta_2(\Delta\mu_2 - \Delta\mu_2^{ub} + t_2^2) \\ & - \varphi_1(\Delta\mu_1 - s_1^2) \\ & - \varphi_2(\Delta\mu_2 - s_2^2), \end{aligned} \tag{4.59}$$

where $\lambda, \theta_1, \theta_2, \varphi_1$, and φ_2 are the Lagrangian multipliers and t_1^2, t_2^2, s_1^2, and s_2^2 are slack variables that are introduced to convert the inequality constrains into equalities

The partial derivatives are:

$$\frac{\partial L}{\partial \Delta\mu_1} = \frac{\partial p_1^*}{\partial \Delta\mu_1}(F_1 + F_2) - \frac{\partial p_1^*}{\partial \Delta\mu_1} p_2^* F_2 - \lambda c_1 - \theta_1 - \varphi_1 \tag{4.60}$$

$$\frac{\partial L}{\partial \Delta\mu_2} = \frac{\partial p_2^*}{\partial \Delta\mu_2}(F_2 + F_3) - \frac{\partial p_2^*}{\partial \Delta\mu_2} p_1^* F_2 - \lambda c_2 - \theta_2 - \varphi_2 \tag{4.61}$$

$$\frac{\partial L}{\partial \lambda} = C - c_1\Delta\mu_1 - c_2\Delta\mu_2 \tag{4.62}$$

$$\frac{\partial L}{\partial \theta_1} = -(\Delta\mu_1 - \Delta\mu_1^{ub} + t_1^2) \tag{4.63}$$

$$\frac{\partial L}{\partial \theta_2} = -(\Delta\mu_2 - \Delta\mu_2^{ub} + t_2^2) \tag{4.64}$$

$$\frac{\partial L}{\partial \varphi_1} = s_1^2 - \Delta\mu_1 \tag{4.65}$$

$$\frac{\partial L}{\partial \varphi_2} = s_2^2 - \Delta\mu_2 \tag{4.66}$$

The optimisation problems with only equality constraints can be solved using the Lagrange multiplier approach, where the partial derivatives of the Lagrangian function are zero at the optimal solution. The complementary slackness condition extends this method to optimization problems with inequality constraints. It states that at the optimal solution, either the corresponding Lagrange multiplier is zero, meaning that the constraint is inactive, or the corresponding slack variable is zero, meaning that the constraint is active. The four inequality constraints mean $2^4 = 16$ cases to discuss for the problem here.

Case 1: None of the inequality constrains is active. $(\theta_1 = 0, t_1^2 > 0), (\theta_2 = 0, t_2^2 > 0), (\varphi_1 = 0, s_1^2 > 0)$, and $(\varphi_2 = 0, s_2^2 > 0)$. The solution can be found by solving:

$$\frac{\partial p_1^*}{\partial \Delta\mu_1}(F_1 + F_2) - \frac{\partial p_1^*}{\partial \Delta\mu_1} p_2^* F_2 - \lambda c_1 = 0 \tag{4.67}$$

$$\frac{\partial p_2^*}{\partial \Delta\mu_2}(F_2 + F_3) - \frac{\partial p_2^*}{\partial \Delta\mu_2} p_1^* F_2 - \lambda c_2 = 0 \tag{4.68}$$

$$C - c_1\Delta\mu_1 - c_2\Delta\mu_2 = 0 \tag{4.69}$$

The two derivatives with respect to $\Delta\mu_1$ and $\Delta\mu_2$ are found as:

$$\frac{\partial p_1^*}{\partial \Delta\mu_1} = -\frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\omega_1-\mu_1-\Delta\mu_1}{\sigma_1})^2} \tag{4.70}$$

$$\frac{\partial p_2^*}{\partial \Delta\mu_2} = -\frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\omega_2-\mu_2-\Delta\mu_2}{\sigma_2})^2} \tag{4.71}$$

With the expressions above, the equations to solve for the optimal solution becomes:

$$-\frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\omega_1-\mu_1-\Delta\mu_1}{\sigma_1})^2}(F_1 + F_2 - p_2^* F_2) - \lambda c_1 = 0 \tag{4.72}$$

$$-\frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\omega_2-\mu_2-\Delta\mu_2}{\sigma_2})^2}(F_2 + F_3 - p_1^* F_2) - \lambda c_2 = 0 \tag{4.73}$$

$$C - c_1\Delta\mu_1 - c_2\Delta\mu_2 = 0 \tag{4.74}$$

The problem becomes difficult to solve because $\Delta\mu_1$ and $\Delta\mu_2$ are part of the power of the natural exponent. Modifications are needed to make the problem analytically solvable.

4.4.3 The Modified Objective Function

To make the optimisation function analytically solvable, instead of maximising the expected delivery, the objective function is modified to maximise the increase in expected delivery benefited from the investment in upgrading the assets. Effectively, the new objective function is the objective function in Equation 4.50 minors the expected delivery without the upgrade. The modification does not change the optimal solution as it simply minor the original objective function E^* by E, which is regard as a constant in this optimisation problem. The objective function to be maximised is therefore:

$$E^* - E = -(p_1^* - p_1)(F_1 + F_2) - (p_2^* - p_2)(F_2 + F_3) + (p_1^* p_2^* - p_1 p_2)F_2, \tag{4.75}$$

or, to minimise,

$$E - E^* = (p_1^* - p_1)(F_1 + F_2)(p_2^* - p_2)(F_2 + F_3) - (p_1^* p_2^* - p_1 p_2)F_2. \tag{4.76}$$

To further simplify the objective function, expressions $p_1^* - p_1$, $p_2^* - p_2$, and $p_1^* p_2^* - p_1 p_2$ are hereby simplified. Given:

$$p_i = \frac{1}{2}\left(1 + \text{erf}\left(\frac{\omega_i - \mu_i}{\sigma_i\sqrt{2}}\right)\right). \tag{4.77}$$

Let $p_i(\mu)$ denotes a function with μ as an input variable:

$$p_i(\mu) = \frac{1}{2}\left(1 + \text{erf}\left(\left(\frac{\omega_i - \mu}{\sigma_i\sqrt{2}}\right)\right)\right). \tag{4.78}$$

Then the derivative of p_i with respect to μ is:

$$p_i' = -\frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\omega_i-\mu}{\sigma_i})^2}, \tag{4.79}$$

which at $\mu = \mu_i$ becomes:

$$p_{\mu_i}' = -\frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\omega_i-\mu_i}{\sigma_i})^2}. \tag{4.80}$$

Let $\Delta\mu_i$ be a sufficiently small increment, the value of $p_i(\mu)$ at $\mu = \mu_i + \Delta\mu_i$ can be approximated by:

$$p_i(\mu = \mu_i + \Delta\mu_i) \simeq p_i + \Delta\mu_i p_{\mu_i}'. \tag{4.81}$$

Under the assumption of $\Delta\mu_1$ and $\Delta\mu_2$ being sufficiently small, expressions $p_1^* - p_1$, $p_2^* - p_2$, and $p_1^* p_2^* - p_1 p_2$ can be simplified to:

$$\begin{aligned} p_1^* - p_1 &= p_1(\mu_1 + \Delta\mu_1) - p_1 \\ &\simeq p_1 + \Delta\mu_1 p_1' - p_i(\mu_1) \\ &= \Delta\mu_1\ p_1', \end{aligned} \tag{4.82}$$

$$\begin{aligned} p_2^* - p_2 &= p_2(\mu_2 + \Delta\mu_2) - p_2 \\ &\simeq p_2 + \Delta\mu_2 p_2' - p_2 \\ &= \Delta\mu_2 p_2', \end{aligned} \tag{4.83}$$

and

$$\begin{aligned} p_1^* p_2^* - p_1 p_2 &= p_1(\mu_1 + \Delta\mu_1) p_2(\mu_2 + \Delta\mu_2) - p_1 p_2 \\ &\simeq \left(p_1 + \Delta\mu_1 p_1'\right)\left(p_2 + \Delta\mu_2 p_2'\right) - p_1 p_2 \\ &= p_1 p_2' \Delta\mu_2 + p_2 p_1' \Delta\mu_1 + p_1' p_2' \Delta\mu_1 \Delta\mu_2. \end{aligned} \tag{4.84}$$

Now the objective function to be minimised, $E - E^*$, becomes:

$$\begin{aligned} E - E^* &= (p_1^* - p_1)(F_1 + F_2) + (p_2^* - p_2)(F_2 + F_3) - (p_1^* p_2^* - p_1 p_2) F_2 \\ &= \Delta\mu_1 p_1'(F_1 + F_2) + \Delta\mu_2 p_2'(F_2 + F_3) - \left(p_1 p_2' \Delta\mu_2 + p_2 p_1' \Delta\mu_1 + p_1' p_2' \Delta\mu_1 \Delta\mu_2\right) F_2. \end{aligned} \tag{4.85}$$

Rearranging Equation 4.85 gives:

$$\begin{aligned} E - E^* = &\left(p_1'(F_1 + F_2) - p_2 p_1' F_2\right) \Delta\mu_1 \\ &+ \left(p_2'(F_2 + F_3) - p_1 p_2' F_2\right) \Delta\mu_2 \\ &- F_2 p_1' p_2' \Delta\mu_1 \Delta\mu_2 \end{aligned} \tag{4.86}$$

To make the objective function easier to handle, some constant terms are replaced with letter A, B, and D (C is taken to represent the total budget). Let

$$A = p_1'(F_1 + F_2) - p_2 p_1' F_2, \tag{4.87}$$
$$B = p_2'(F_2 + F_3) - p_1 p_2' F_2, \tag{4.88}$$
$$D = -F_2 p_1' p_2'. \tag{4.89}$$

Equation 4.86 becomes:

$$E - E^* = A\Delta\mu_1 + B\Delta\mu_2 + D\Delta\mu_1\Delta\mu_2. \tag{4.90}$$

The optimisation problem now states:

$$\min\left(A\Delta\mu_1 + B\Delta\mu_2 + D\Delta\mu_1\Delta\mu_2\right), \tag{4.91}$$

subject to Equation 4.52 - 4.56.

The Lagrangian function is defined as:

$$\begin{aligned} L(\Delta\mu_1, \Delta\mu_2, \lambda, \theta_1, \theta_2, \varphi_1, \varphi_2) = {} & A\Delta\mu_1 + B\Delta\mu_2 + D\Delta\mu_1\Delta\mu_2 \\ & - \lambda(c_1\Delta\mu_1 + c_2\Delta\mu_2 - m(c_1 + c_2)) \\ & - \theta_1(\Delta\mu_1 - \Delta\mu_1^{ub} + t_1^2) \\ & - \theta_2(\Delta\mu_2 - \Delta\mu_2^{ub} + t_2^2) \\ & - \varphi_1(\Delta\mu_1 - s_1^2) \\ & - \varphi_2(\Delta\mu_2 - s_2^2). \end{aligned} \tag{4.92}$$

With the gradients being:

$$\frac{\partial L}{\partial \Delta\mu_1} = A + D\Delta\mu_2, -\lambda c_1 - \theta_1 - \varphi_1, \tag{4.93}$$
$$\frac{\partial L}{\partial \Delta\mu_2} = B + D\Delta\mu_1 - \lambda c_2 - \theta_2 - \varphi_2, \tag{4.94}$$
$$\frac{\partial L}{\partial \lambda} = C - c_1\Delta\mu_1 - c_2\Delta\mu_2, \tag{4.95}$$
$$\frac{\partial L}{\partial \theta_1} = -(\Delta\mu_1 - \Delta\mu_1^{ub} + t_1^2), \tag{4.96}$$
$$\frac{\partial L}{\partial \theta_2} = -(\Delta\mu_2 - \Delta\mu_2^{ub} + t_2^2), \tag{4.97}$$
$$\frac{\partial L}{\partial \varphi_1} = s_1^2 - \Delta\mu_1, \tag{4.98}$$
$$\frac{\partial L}{\partial \varphi_2} = s_2^2 - \Delta\mu_2, \tag{4.99}$$

where $\lambda, \theta_1, \theta_2, \varphi_1$, and φ_2 are the Lagrangian multipliers and t_1^2, t_2^2, s_1^2, and s_2^2 are the slack

variables that are introduced to convert the inequality constrains into equalities. Again, the four inequality constraints mean $2^4 = 16$ cases to discuss. The solution found in each case could be a local minima. Therefore, for each case, the feasibility of the solved solution is determined by checking if it satisfies the assumptions of the case. Feasible solutions from all cases are compared at the end to determine which feasible solution is the optimal solution.

Case 1: None of the inequality constrains is active. $(\theta_1 = 0, t_1^2 > 0), (\theta_2 = 0, t_2^2 > 0), (\varphi_1 = 0, s_1^2 > 0)$, and $(\varphi_2 = 0, s_2^2 > 0)$. The solution can be found by solving:

$$\begin{cases} A + D\Delta\mu_2 - \lambda c_1 = 0 \\ B + D\Delta\mu_1 - \lambda c_2 = 0 \\ C - c_1\Delta\mu_1 - c_2\Delta\mu_2 = 0 \end{cases} \tag{4.100}$$

The solution is:

$$\Delta\mu_1 = \frac{CD + Ac_2 - Bc_1}{2Dc_1} \tag{4.101}$$

$$\Delta\mu_2 = \frac{CD - Ac_2 + Bc_1}{2Dc_2} \tag{4.102}$$

Feasibility of the above solution need to be checked with if there exist $t_1^2 > 0, t_2^2 > 0, s_1^2 > 0$, and $s_2^2 > 0$ that satisfy:

$$\frac{\partial L}{\partial \theta_1} = \frac{\partial L}{\partial \theta_2} = \frac{\partial L}{\partial \varphi_1} = \frac{\partial L}{\partial \varphi_2} = 0. \tag{4.103}$$

If any of the gradients can not find a slack variable so that the gradient can be zero, the solution from this case will be ruled unfeasible. This situation happens when the solution found in this case turns out to be outside the feasible region defined by the constraints. When there are no constraints considered, the solution calculated with Equation 4.101 and Equation 4.102 is the optimal solution. With the presence of the equality and inequality constraints, it could turn out to be an unfeasible solution.

Case 2: The first inequality constrain is active. $(\theta_1 \neq 0, t_1^2 = 0), (\theta_2 = 0, t_2^2 > 0), (\varphi_1 = 0, s_1^2 > 0)$, and $(\varphi_2 = 0, s_2^2 > 0)$. By solving

$$\frac{\partial L}{\partial \theta_1} = 0 \quad \text{and} \quad \frac{\partial L}{\partial \lambda} = 0, \tag{4.104}$$

the solution is

$$\begin{aligned} \Delta\mu_1 &= \Delta\mu_1^{ub} \\ \Delta\mu_2 &= \frac{C - c_1\Delta\mu_1^{ub}}{c_2} \end{aligned} \tag{4.105}$$

In this case, available resources are firstly used to saturate the upgrade of edge e_1 until reaching the upper limit. Remaining resources is then used to upgrade edge e_2. Feasibility of this solution need to satisfy $\Delta\mu_2 > 0$.

Case 3: The second inequality constrain is active. $(\theta_1 = 0, t_1^2 > 0), (\theta_2 \neq 0, t_2^2 = 0), (\varphi_1 = 0, s_1^2 > 0)$, and $(\varphi_2 = 0, s_2^2 > 0)$. By solving

$$\frac{\partial L}{\partial \lambda} = 0 \quad \text{and} \quad \frac{\partial L}{\partial \theta_2} = 0 \tag{4.106}$$

the solution is:

$$\begin{aligned} \Delta\mu_1 &= \frac{C - c_2 \Delta\mu_2^{ub}}{c_1}, \\ \Delta\mu_2 &= \Delta\mu_2^{ub}. \end{aligned} \tag{4.107}$$

Feasibility of this solution need to satisfy $\Delta\mu_1 > 0$.

Case 4: The third inequality constrain is active. In this case, edge e_1 have no resources allocated and edge e_2 has all of it. The solution is:

$$\begin{aligned} \Delta\mu_1 &= 0 \\ \Delta\mu_2 &= C/c_2. \end{aligned} \tag{4.108}$$

Feasibility of this solution need to satisfy $C/c_2 < \Delta\mu_2^{ub}$.

Case 5: The forth inequality constrain is active. In this case, edge e_2 have no resources allocated and edge e_1 has all of it. The solution is:

$$\begin{aligned} \Delta\mu_1 &= C/c_1 \\ \Delta\mu_2 &= 0. \end{aligned} \tag{4.109}$$

Feasibility of this solution need to satisfy $C/c_1 < \Delta\mu_1^{ub}$.

Case 6: The first and fourth inequality constrains are active. $(\theta_1 \neq 0, t_1^2 = 0), (\theta_2 = 0, t_2^2 > 0), (\varphi_1 = 0, s_1^2 > 0)$, and $(\varphi_2 \neq 0, s_2^2 = 0)$. The solution is:

$$\begin{aligned} \Delta\mu_1 &= \Delta\mu_1^{ub} \\ \Delta\mu_2 &= 0. \end{aligned} \tag{4.110}$$

Feasibility of this case need to satisfy $c_1 \Delta\mu_1^{ub} = C$.

Case 7: The second and third inequality constrains are active. $(\theta_1 = 0, t_1^2 > 0), (\theta_2 \neq 0, t_2^2 = 0), (\varphi_1 \neq 0, s_1^2 = 0)$, and $(\varphi_2 = 0, s_2^2 > 0)$. Like Case 5, the solution is:

$$\begin{aligned} \Delta\mu_1 &= 0 \\ \Delta\mu_2 &= \Delta\mu_2^{ub}. \end{aligned} \tag{4.111}$$

Feasibility of this case need to satisfy $c_2 \Delta\mu_2^{ub} = C$.

Case 8: The first and second inequality constrains are active. $(\theta_1 \neq 0, t_1^2 = 0), (\theta_2 \neq$

$0, t_2^2 = 0), (\varphi_1 = 0, s_1^2 > 0)$, and $(\varphi_2 = 0, s_2^2 > 0)$. The solution is simply:

$$\begin{aligned} \Delta\mu_1 &= \Delta\mu_1^{ub} \\ \Delta\mu_2 &= \Delta\mu_2^{ub}. \end{aligned} \tag{4.112}$$

Feasibility of this case need to satisfy $c_1\Delta\mu_1^{ub} + c_2\Delta\mu_2^{ub} = C$.

Case 9: The first and third inequality constrain are active. This is not a realistic case as $\Delta\mu_1$ can not equal to 0 and $\Delta\mu_1^{ub}$ at the same time unless $\Delta\mu_1^{ub} = 0$.

Case 10: The first, third and forth inequality constrains are active. Unrealistic for the same reason as Case 9.

Case 11: The second and forth inequality constrains are active. Unrealistic case as $\Delta\mu_2$ can not equal to 0 and $\Delta\mu_2^{ub}$ at the same time unless $\Delta\mu_2^{ub} = 0$.

Case 12: The second, third and forth inequality constrains are active. Unrealistic for the same reason as Case 11.

Case 13: The third and forth inequality constrains are active. This is not possible unless there is no resource input and $C = 0$.

Case 14: The first, second and third inequality constrains are active. Unrealistic case as $\Delta\mu_1$ can not equal to 0 and $\Delta\mu_1^{ub}$ at the same time unless $\Delta\mu_1^{ub} = 0$.

Case 15: The first, second and third inequality constrains are active. Unrealistic case as $\Delta\mu_2$ can not equal to 0 and $\Delta\mu_2^{ub}$ at the same time unless $\Delta\mu_2^{ub} = 0$.

Case 16: All inequality constrains are active. Unrealistic case.

With the solutions from all 16 cases in analytical form, it is not easy to evaluate the feasibility and optimum of the 16 solutions without numerical values. This is as far as the Lagrange multiplier method can achieve.

4.5 The Unified Approach

Among the 16 solutions, the case 1 solution is the optimal solution if there is no lower or upper boundary. Some further simplifications are made to Equation 4.101 and 4.102 in order to help better understand the relationship between $\Delta\mu_i$ and the variables. If the two edges of the same unit cost, with $(c_1, c_2) = (c, c)$:

$$\Delta\mu_1 = \frac{C}{2c} + \frac{A - B}{2D}, \tag{4.113}$$

and

$$\Delta\mu_2 = \frac{C}{2c} + \frac{B - A}{2D}. \tag{4.114}$$

With the above simplification, it is noticed that the solution is related to the relative values of A and B, which can be arranged as:

$$A = p_1'(F_1 + F_2) - p_2 p_1' F_2 = p_1'(F_1 + (1 - p_2)F_2),$$

and

$$B = p_2'(F_2 + F_3) - p_1 p_2' F_2 = p_2'(F_3 + (1 - p_1)F_2).$$

A is interpreted as the changing rate of its probability of failure, p_1, multiplied by the amount of flow that is dependent on the success of edge e_1. As $F1$ in this simple network model relies solely on the success of edge e_1, therefore $F1$ is attributed to edge e_1 with no other multiplications. Unlike $F1$, $F2$ relies on both e_1 and e_2 and the expected delivery is affected by the probabilities of failure of both edges. Therefore, when attributing $F2$ to edge e_1, to quantify the effect of investing in edge e_1, $F2$ is multiplied by edge e_2's probability of success, $(1 - p_2)$.

With this understanding established, putting (c_1, c_2) back into Equation 4.101 and rearranging gives:

$$\Delta\mu_1 = \frac{CD + Ac_2 - Bc_1}{2Dc_1}$$

$$\Rightarrow \; c_1\Delta\mu_1 - \frac{C}{2} = \frac{c_1 c_2}{D}\left(\frac{A}{c_1} - \frac{B}{c_2}\right)$$

$$\Rightarrow \; c_1\Delta\mu_1 - \frac{C}{2} = \frac{2c_1 c_2}{D}\left(\frac{A}{c_1} - \frac{1}{2}\left(\frac{A}{c_1} + \frac{B}{c_2}\right)\right). \tag{4.115}$$

Similarly, for Equation 4.102, rearranging gives:

$$c_2\Delta\mu_2 - \frac{C}{2} = \frac{2c_1 c_2}{D}\left(\frac{B}{c_2} - \frac{1}{2}\left(\frac{A}{c_1} + \frac{B}{c_2}\right)\right). \tag{4.116}$$

From Equation 4.115, in the optimised resource allocation solution, the amount of more-than-average resource allocated to edge e_1 is a linear function of by how much A/c_1 is more than the average of $(A/c_1, B/c_2)$. The $(F_1 + (1 - p_2)F_2)$ in A and the $(F_3 + (1 - p_1)F_2)$ in B need to be generalised so that it can be applied to problems with more than two edges. Start with edge $e_{i=1}$, the direct sum of all OD pairs that have edge e_1 on its path, without considering the probability of failures, is:

$$\sum_j H_{1,j} F_j$$

Introducing the probability of failure to calculate the true proportion of F_j that contributes to the ranking of edge $e_{i=1}$, is:

$$\prod_{k,k\neq 1} (1 - H_{k,j} p_k).$$

$A = p_1'(F_1 + (1 - p_2)F_2)$ and $B = p_2'(F_3 + (1 - p_1)F_2)$ are hence generalised to:

$$p_i' \sum_j H_{i,j} F_j \prod_{k,k \neq i} (1 - H_{k,j} p_k). \tag{4.117}$$

Generalised A/c_1 and B/c_2 are therefore:

$$\frac{p_i'}{c_i} \sum_j H_{i,j} F_j \prod_{k,k \neq i} (1 - H_{k,j} p_k). \tag{4.118}$$

Among the three candidate network metrics developed in Section 4.3, the gradient metric shows a degree of similarity with the above equation. To recall, the gradient of investing in edge e_i is:

$$\frac{\partial E}{\partial \mu_i} = -\frac{\partial p_i}{\partial \mu_i} \sum_j H_{i,j} F_j \prod_{k,k \neq i} (1 - H_{k,j} p_k). \tag{4.119}$$

There is a striking similarity between the metric developed (Equation 4.119)and the generalised analytical solution obtained (Equation 4.118). The term $\partial p_i / \partial \mu_i$ in Equation 4.119 is the same as the term p_i' in Equation 4.118. The difference is that the network metric does not factor in the unit cost.

The term $\frac{\partial E}{\partial \mu_i}$ indicates the gradient of investing in edge e_i. Assuming there is a sufficiently small portion, $\Delta\mu_i$ allocated to edge e_i, the resulted change in the objective function ΔE can be approximated as:

$$\Delta E = \Delta \mu_i \frac{\partial E}{\partial \mu_i}. \tag{4.120}$$

With the associated cost of upgrading edge e_i by $\Delta\mu_i$ being $\Delta\mu_i c_i$, the true gradient of investing in edge e_i is:

$$\frac{\Delta E}{\Delta \mu_i c_i} = \frac{\Delta \mu_i \frac{\partial E}{\partial \mu_i}}{\Delta \mu_i c_i} = -\frac{p_i'}{c_i} \sum_j H_{i,j} F_j \prod_{k,k \neq i} (1 - H_{k,j} p_k). \tag{4.121}$$

Now the third candidate network metric developed meets the generalised analytical solution Equation 4.118, pointing to the same formulation. From the consistency between the solution arising from the two completely different approaches, it can be concluded that a metric-based resource allocation strategy using the identified gradient has the potential to allocate resources effectively towards climate resilience of infrastructure systems. This analytic derivation will be applied and demonstrated in the next chapter.

4.6 Chapter Conclusion

To answer the research question of how should resources be allocated in the most efficient manner to upgrade infrastructure assets so that the forthcoming service disruptions can be minimised, this chapter considers the problem of finding the optimal or best possible resource allocation

between infrastructure assets to minimise the system's service disruptions under plausible future extreme weather events with a finite budget.

The resource allocation problem is formulated as a mathematical optimisation problem that takes the future weather pattern as model inputs and aims to maximise the climate resilience of the system. Using the metric-based resource allocation method, a metric is derived from the formulated objective function. The 'gradient' metric captures the topological structure of the network, supply-demand features, and future weather failure patterns. The metric can be used to allocate resources among a large number of assets in a computationally tractable manner as described in Section 4.3. Using a simple network model with 2 assets, an analytical solution is obtained using the Lagrangian multiplier method. The two streams of work meet each other as it is observed that the gradient metric and the analytic solution from the simple system model converge to the same expression. In this respect, a ranking strategy using the observed metric could potentially be able to allocate the resources efficiently, eliminating the need to solve those high-dimensional non-linear programming problems.

Having identified the common expression in this chapter, the following chapter presents the application of the network-metric-based resource allocation method with three case studies, where the proposed metric is further developed, tested and confirmed its usefulness in efficiently directing resource distribution towards climate resilience of infrastructure systems.

Chapter 5

Case Studies For Asset Upgrade

This chapter presents three case studies. Case study 1 uses the the same test model as the previous chapter to compare the performance of 2 numerical optimisation algorithm (a Sequential Quadratic Programming (SQP) method and an interior-point method), the Lagrangian multiplier method as solved in Chapter 4 and the proposed metric-based method with the 'gradient' metric. Using 10,000 randomly configured values for the system model, the metric-based method is shown to be nearly as good as the numerical method. It was also found that the metric-based method works better with smaller amount of resources, leading to the development of an iterative metric-based resource allocation approach.

Case study 2 uses a group of synthetic network models with the asset layer comprising 3 to 6 nodes to compare the performance of a numerical method and the proposed metric-based method. 10,500 system models were randomly configured and 42,000 optimisation cases were solved. It was found that the percentage difference between the numerical solution and the metric-based method has a trend to plateau, meaning the metric-based method scales well with increasing dimensions unlike normal solution-finding algorithms. Using the same system configurations, this case study also tested the iterative metric-based approach and found that by iterating the metric-based method several times, the solution found can approach the numerical solution very well.

Case study 3 applies the developed iterative metric-based method to a core rail freight network model of Great Britain with 84 edges in the asset layer. The performance of the proposed method is benchmarked with the Particle Swarm Optimisation (PSO) algorithm. The simulation shows that for this 84-dimensional problem, the iterative metric-based method can yield better solutions using significantly less computational resources. This case study also provides some important insights. Inequality is associated with resource scarcity. When resources are limited it may be more beneficial to invest in fewer but more critical assets.

The network models in Case study 1 and Case study 2 are regard as smaller scale synthetic models of the core freight network model of the case study 3. In those small scale models, the edges in the asset layer represents railway track segments and the flows are freight transportation.

5.1 Case Study 1 - A Simple Network Model

This case study approaches the same 2D optimisation problem as formulated in Chapter 4 Section 4.4. Three groups of different methods are used to solve the same 2D optimisation problem in Figure 5.1:

- The metric-based resource allocation method with the gradient metric identified,
- The Lagrangian multiplier method,
- Two numerical methods, namely a SQP method and an interior-point method.

5.1.1 The Network Model and the Optimisation Problem

The asset layer in the simple network model has 3 nodes and 2 edges as shown in Figure 5.1. The current design loads are denoted as μ_1 and μ_2 for edge e_1 and e_2. The service layer comprises three OD pairs, denoted as F_1, F_2, and F_3. The route-link incidence matrix, H, is:

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

5.1.2 The Optimisation Problem

The optimisation problem is to find the optimal allocation of a given budget (C) that minimises the initial loss of services (iLOS) under the given external weather parameters (ω_1, ω_2) with to the specified unit costs (c_1, c_2) of upgrading edges, constrained by the given upper boundaries $(\Delta\mu_1^{ub}, \Delta\mu_2^{ub})$. In Section 4.4.3, the problem was formulated:

$$\min_{\boldsymbol{\Delta\mu}} \text{iLOS} = p_1^*(F_1 + F_2) + p_2^*(F_2 + F_3) - p_1^* p_2^* F_2, \tag{5.1}$$

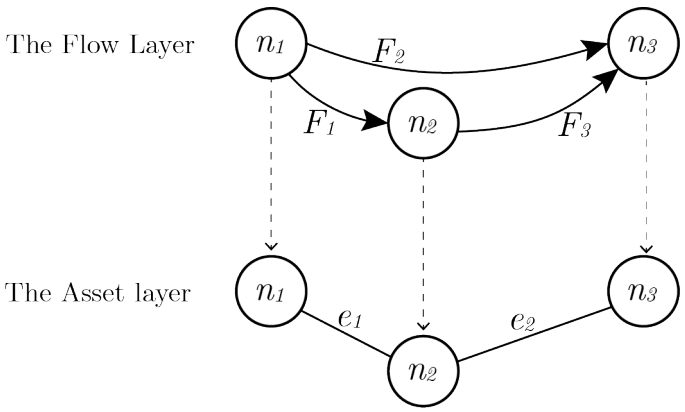


Figure 5.1: The network structure of the testing model. The flow layer has 3 OD pairs. The asset layer is a network of 3 nodes and 3 edges.

subject to:

$$c_1\Delta\mu_1 + c_2\Delta\mu_2 = C, \tag{5.2}$$

$$\Delta\mu_1 \leq \Delta\mu_1^{ub}, \tag{5.3}$$

$$\Delta\mu_2 \leq \Delta\mu_2^{ub}, \tag{5.4}$$

$$\Delta\mu_1 \geq 0, \tag{5.5}$$

$$\Delta\mu_2 \geq 0, \tag{5.6}$$

where,

$$p_1^* = \frac{1}{2}\left[1 + \text{erf}\left(\frac{\omega_1 - (\mu_1 + \Delta\mu_1)}{\sigma_1\sqrt{2}}\right)\right], \tag{5.7}$$

$$p_2^* = \frac{1}{2}\left[1 + \text{erf}\left(\frac{\omega_2 - (\mu_2 + \Delta\mu_2)}{\sigma_2\sqrt{2}}\right)\right]. \tag{5.8}$$

In order to apply the numerical methods for the optimisation problem, the values of the weather parameters (ω_1, ω_2), the OD flows (F_1, F_2, F_3), and the edges' unit cost (c_1, c_2) are independently and randomly sampled using three random variables that are of continuous uniform distributions between the bounds:

$$\omega_1, \omega_2 \leftarrow U(30, 40), \tag{5.9}$$

$$F_1, F_2, F_3 \leftarrow U(1, 100), \tag{5.10}$$

$$c_1, c_2 \leftarrow U(1, 00). \tag{5.11}$$

Together the value of design load and external weather parameter determine the probability of failure of an edge. Low design load with low external temperature could return a probability of failure that is as high as one with high external temperature with high design load, as shown in Equation 5.12.

$$\frac{1}{2}\left[1 + \text{erf}\left(\frac{\omega - (\mu + \Delta\mu)}{\sigma\sqrt{2}}\right)\right] = \frac{1}{2}\left[1 + \text{erf}\left(\frac{\omega + x - (\mu + x + \Delta\mu)}{\sigma\sqrt{2}}\right)\right]. \tag{5.12}$$

It means an edge of design load μ subjected to external condition ω returns the same probability of failure as an edge of design load $\mu + x$ subjected to external condition $\omega + x$. In order to cover as many combinations as possible for a fixed number of sampling, the edge's design load $(\mu_1,\ \mu_2)$ are assumed to be 35°C, and only the values of the weather parameter ω_1 and ω_2 are randomly sampled. After the values for those input variables are generated, the flow vector $F = (F_1, F_2, F_3)$ is then normalised so that the sum always equals to 100. In this way, the expected delivery when the system suffers zero damage is always 100, which makes the results comparable between the sampled flow vectors. The cost vector $\boldsymbol{c} = (c_1, c_2)$ is also normalised so that the sum is always 1.

The term *a system configuration* will be used in this chapter to refer to a single instance of the system model with a given set of sampled $H, F, \boldsymbol{\omega}$, and $\boldsymbol{c}$. Each system configuration is subjected

to various levels of upgrading resources C. The overall level of resource input is quantified with a separate variable, m, which is defined as $C = m\sum_i c_i$. A resource input of $m\sum_i c_i$ means the invested resources is enough to upgrade all asset in the network by m°C. The values of m can be related to an adaptation measure in response to global temperature increases. When m is low, the input resource is scarce, while a high value of m indicates a generous input. When formulating the network metric in Chapter 4, Section 4.5, it was assumed that $\Delta\mu_i$ is sufficiently small so that the increment in the objective function can be approximated with the local derivative. To minimise the possible violation of this assumption, the resource sufficiency indicator in this case study is set to four discrete values as $m \in [0.25, 0.5, 0.75, 1]$.

5.1.3 Solving the Optimisation Problem

Before presenting the solutions, the expressions of a few terms are firstly presented. The probabilities of failure for the edges in its current condition before the upgrades are:

$$p_1 = \frac{1}{2}\left(1 + \text{erf}\left(\frac{\omega - \mu_1}{\sigma_1\sqrt{2}}\right)\right), \tag{5.13}$$

$$p_2 = \frac{1}{2}\left(1 + \text{erf}\left(\frac{\omega - \mu_2}{\sigma_2\sqrt{2}}\right)\right). \tag{5.14}$$

The local derivates of the probability of failure with respect to μ at $\mu = \mu_1$, and $\mu = \mu_2$ are expressed as:

$$p_1' = -\frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\omega_1 - \mu_1}{\sigma_1})^2}, \tag{5.15}$$

$$p_2' = -\frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\omega_2 - \mu_2}{\sigma_2})^2}. \tag{5.16}$$

The Metric-based Method

As defined in Chapter 4 Section 4.4.3, the solution from the metric-based resource allocation is calculated as:

$$\Delta\mu_i = \gamma_i \frac{C}{\sum_{i=1} c_i\gamma_i}, \tag{5.17}$$

where γ_i is the calculated metric for edge e_i. For the network model here, the metrics are:

$$\gamma_1 = p_1'(F_1 + F_2 - p_2F_2)/c_1, \tag{5.18}$$

$$\gamma_2 = p_2'(F_3 + F_2 - p_1F_2)/c_2. \tag{5.19}$$

Therefore, the solutions are:

$$\Delta\mu_1 = \frac{Cp_1'(F_1 + (1-p_2)F_2)}{c_1\left(p_1'(F_1+F_2-p_2F_2) + p_2'(F_3+F_2-p_1F_2)\right)}, \tag{5.20}$$

$$\Delta\mu_i = \frac{Cp_2'(F_3 + F_2 - p_1F_2)}{c_2\left(p_1'(F_1+F_2-p_2F_2) + p_2'(F_3+F_2-p_1F_2)\right)}. \tag{5.21}$$

The metric-based resource allocation method naturally gives solutions where $\Delta\mu_i$ are greater than zero and satisfies the equality constraint $c_1\Delta\mu_1 + c_2\Delta\mu_2 = C$. However, the inequality constraint for the upper boundary is difficult to enforce. To keep things simple, the upper boundary constraint is temporarily lifted in this case study and will be re-introduced in Case study 2 and 3. Accordingly, the numerical method and the Lagrangian method in this case study both have a lower boundary of 0 and no upper boundary.

The Lagrangian Multiplier Method

With the upper boundary constraint lifted, the number of cases to discuss reduces from 16 to 4. With the alternative objective function (Equation 4.76), the solution obtained from the case where none of the inequality constrains is active is:

$$\Delta\mu_1 = \frac{CD + Ac_2 - Bc_1}{2Dc_1}, \tag{5.22}$$

$$\Delta\mu_2 = \frac{CD - Ac_2 + Bc_1}{2Dc_2}, \tag{5.23}$$

where,

$$A = p_1'(F_1 + F_2) - p_2p_1'F_2, \tag{5.24}$$

$$B = p_2'(F_2 + F_3) - p_1p_2'F_2, \tag{5.25}$$

$$D = -F_2p_1'p_2'. \tag{5.26}$$

The feasibility of the solution is determined by checking if the $\Delta\mu_1$ and $\Delta\mu_2$ calculated with Equation 5.22 and Equation 5.23 are greater than 0. The second case is that the inequality constrain $\Delta\mu_1 \geq 0$ is active. In this case, edge e_1 has no resources allocated and edge e_2 has all of the resource input. The solution is:

$$\Delta\mu_1 = 0, \tag{5.27}$$

$$\Delta\mu_2 = C/c_2. \tag{5.28}$$

The third case is that the inequality constrain $\Delta\mu_2 \geq 0$ is active. In this case, edge e_2 has no resources allocated and edge e_1 has all of the resource input. The solution is:

$$\Delta\mu_1 = C/c_1, \tag{5.29}$$

$$\Delta\mu_2 = 0. \tag{5.30}$$

The last case is that both inequality constraints are active where $\Delta\mu_1 = \Delta\mu_2 = 0$, which does not satisfy the cost constraint and is therefore infeasible. For a given system configuration and a given amount of resource input, feasible solutions from the three cases are compared, and the solution that gives the lowest objective function value is the optimal solution.

Numerical Techniques

The characteristics of the optimisation problem are the driving factors in determining what solution-finding algorithms can be applied. The objective function formulated, Equation 5.1, contains the fragility function. The key characteristics of the fragility function are firstly discussed below.

- **Non-linear** - Figure 5.2 shows a plot of Equation 5.7 with $\omega_1 = 45, \mu_1 = 35, \sigma_1 = 2.5$ as an example. As shown in the figure, the probability of failure, decreases with increasing $\Delta\mu_1$. The rate of change (the gradient of the curves) is relatively small with small $\Delta\mu_1$ but gradually increases before decrease again.

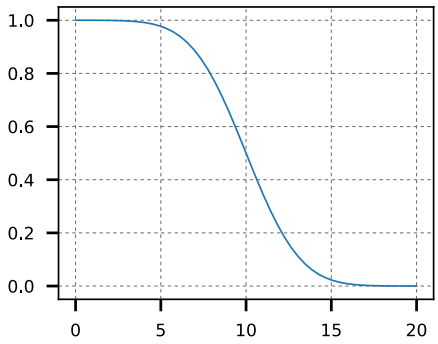


Figure 5.2: Equation 5.7, the fragility function with $\omega_1 = 45, \mu_1 = 35, \sigma_1 = 2.5$. The x-axis is the value of $\Delta\mu_1$ and the y-axis is the probability of failure.

- **Non-convex** - The concept of convexity is fundamental in optimisation. If the objective function is convex in its feasible region, the problem is much easier to solve because the local minima found will be the global minima (Nocedal and Wright, 2006). Otherwise, there may exist multiple local minima and any local minima found can not be guaranteed to be the global minima. Drawing a line from $(x = 0, y = p_1(0))$ to point $x = 20, y = p_1(20))$ in Figure 5.2 shows that Equation 5.7 is not a convex function and therefore the objective function is not a convex function.

• **Monotonic** - The probability of failure for a single edge is always decreasing with increasing $\Delta\mu_i$. The first derivative of the objective function with respect to $\Delta\mu_i$ is always negative.

• **Differentiability** - The fragility function is twice differentiable as its first derivatives, Equation 5.15 and Equation 5.16, are continuous.

Commonly used numerical methods for multivariate non-linear optimisation with both equality and inequality constraints can be broadly grouped into two categories - direct search and gradient-based method. The main difference is that the direct search method does not require the derivatives of the objective function in searching for the optimal solution, while the gradient-based method does. The gradient method is preferred as the objective function here is twice differentiable. According to the classification in Nocedal and Wright (2006), options for the problem here are:

- augmented Lagrangian methods,
- sequential quadratic programming (SQP) method,
- interior-point methods / barrier method.

SciPy provides Python implementations of the latter two methods, which are therefore chosen as the numerical method for this case study. SciPy's minimisation method, Sequential Least SQuares Programming (SLSQP), applies to general non-linear optimisation problems with both equality and inequality constraints. SciPy's minimisation method *trust-constr* uses a trust-region interior point algorithm for constrained optimisation with equality and inequality constraints. Both of the two methods are used for this case study.

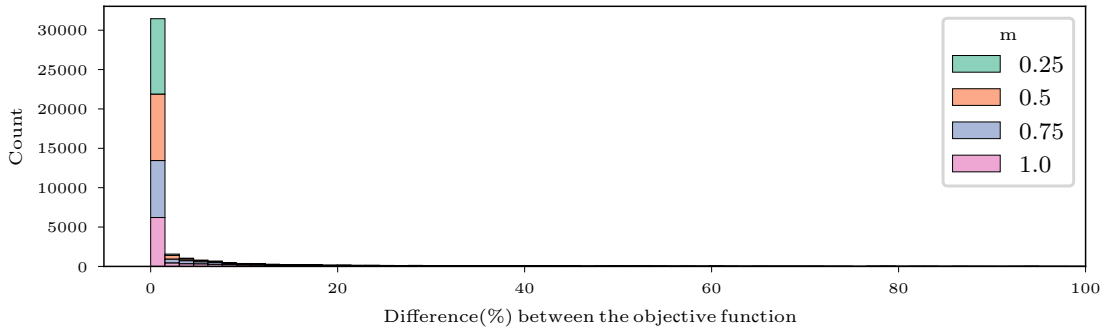
5.1.4 Results

The values of the objective functions calculated with the solutions found using the four methods are compared. The names of the four methods are shortened as:

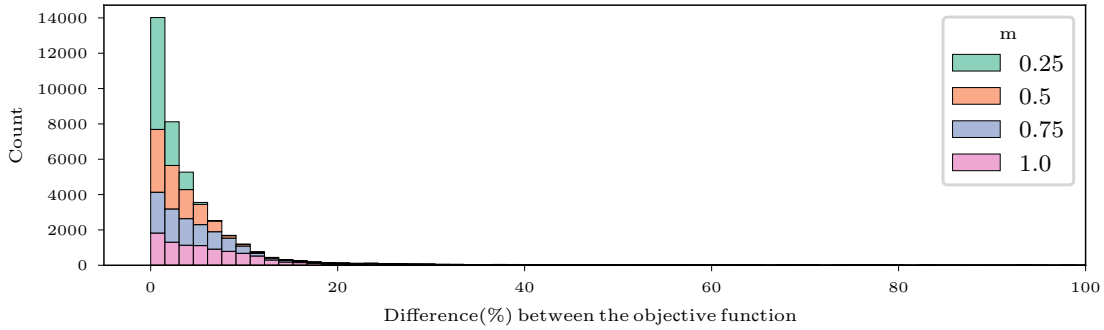
- The metric-based resource allocation method ⇒ metric-based
- The Lagrangian multiplier method ⇒ L-multiplier
- The first numerical method - SQP ⇒ SQP
- The second numerical method - the interior-point method ⇒ interior-point

SQP v.s. interior-point

The two numerical methods, namely SQP and the interior-point method, return solutions that are virtually the same. The average percentage difference from 5000 randomly generated system configurations is 0.004%. The two numerical methods cross-validated the optimum of each other's solution. The solutions obtained with the two methods are therefore considered as the optimal solution for this 2D optimisation problem. It should not be a surprise that one of the



(a) Numerical v.s. Lagrange multiplier



(b) Numerical v.s. metric-based

Figure 5.3: Histogram of percentage difference between the objective function with solutions obtained from the three different methods. The comparisons are made between the values of the objective function and calculated as percentage difference.

most widely-used programming libraries can solve a 2D optimisation problem. In the following parts of this subsection, the solution obtained with the SQP will be used as *the numerical solution.*

SQP v.s. L-multiplier v.s. metric-based

To compare the three methods, namely the numerical method (SQP), the Lagrangian multiplier method and the metric-based method, 10000 system configurations are generated. The three methods are applied to each system configuration to find the optimisation solutions for all four investment scenarios, $m \in [0.25, 0.5, 0.75, 1]$. For each combination of a system configuration and an investment scenario, comparisons are made between the three minimised values of the objective function obtained with the three methods. Considering the numerical method's solution is very likely to be the global minimum for this 2D problem, the numerical solutions, therefore, are regarded as the optimal solutions and the comparison baseline. Hence, the percentage difference is calculated as the absolute difference between the two objective functions divided by that of the numerical method.

To check the sanity of the Lagrangian multiplier solutions (Equation 5.22 and Equation 5.23), Figure 5.3 plots the distribution of the percentage differences calculated. Comparing the Lagrangian method to the numerical solution, the percentage difference from the 10,000 system

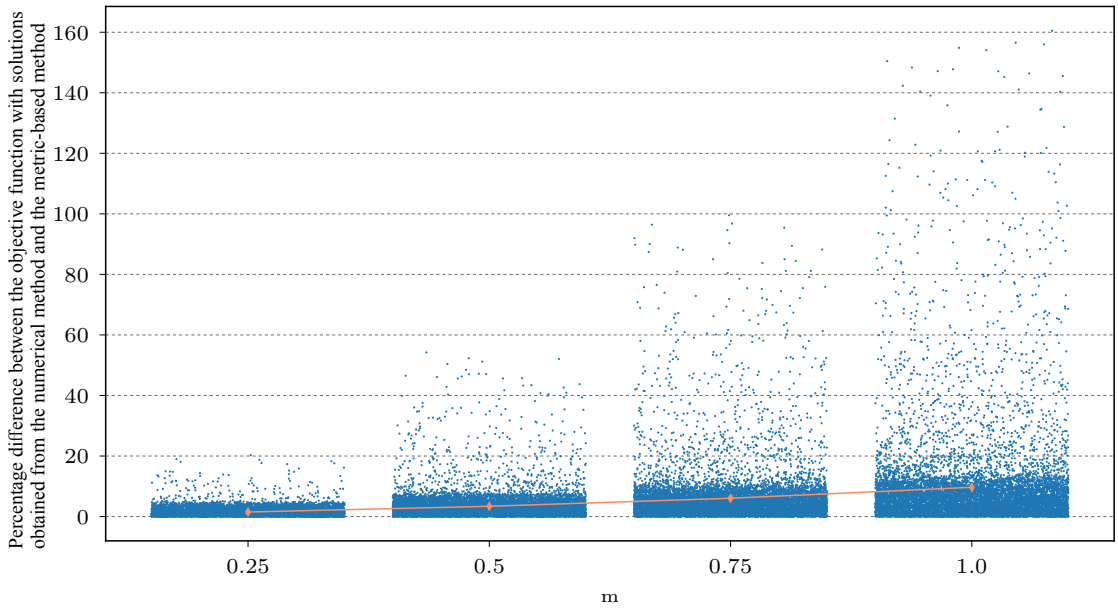
configurations averages at 3.75%. From Figure 5.3a, the distribution is heavily skewed. The median is 2.5E-13%. The small percentage differences confirms that the Lagrangian multiplier solutions (Equation 5.22 and Equation 5.23) agrees with the solutions found by the numerical method.

Figure 5.3b plots the comparison between the metric-based method and the numerical method. The average percentage difference between the two methods from the 10,000 system configurations is 5.06%. The median is 2.5%, meaning half of the times the metric-based method can find a solution that is nearly as optimal as the numerical method. The 90th percentile is 10.13%, meaning for about 90% of the times, the metric-based method can minimise the objective function within 10% of that minimised using the numerical method. It can therefore conclude that the metric-based method using the identified gradient metric works well for this simple system model.

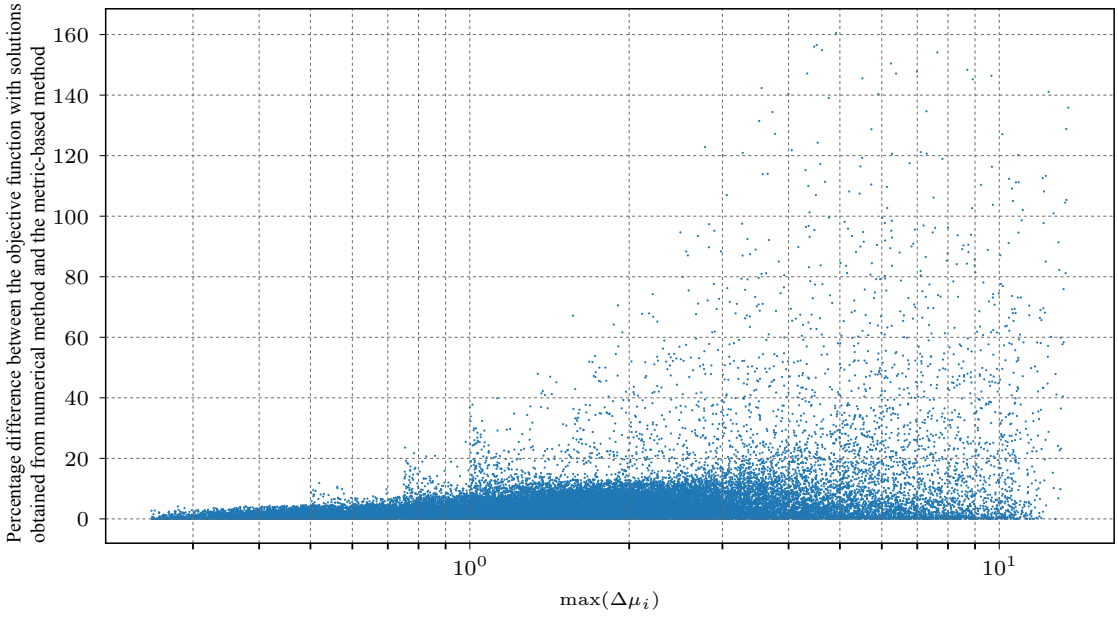
5.1.5 Discussion

From Figure 5.4a, the percentage difference between the objective functions of the metric-based method and the numerical method increases with m. A possible explanation for this increase is that the metric-based method uses the local derivatives to approximate the objective function. An increasing value of m is more likely to result in an edge having a $\Delta\mu_i$ that no longer satisfies the 'sufficiently small' assumption. When $\Delta\mu_i$ is larger than 'sufficiently small', the gradient may change significantly and the change in the objective function is no longer close to the ΔE estimated with the gradient.

Figure 5.4b plots the percentage difference against the larger one between $\Delta\mu_1$ and $\Delta\mu_2$. A large number of dots clusters at the bottom of the plot. Towards the right-hand side of the plot, the dots diffuse more to the top of the plot, meaning a large percentage difference is more likely to occur when the solution generated contains a high value of $\Delta\mu_i$. It further confirms the hypothesis that the magnitude of $\Delta\mu_i$ affects the magnitude of the percentage difference. This understanding leads to the idea of further developing the metric-based strategy to an iterative one. The resources are divided into smaller portions and in each iteration, the metric-based method is used to allocated a single portion to assets. By allocating a portion of resources and update the metrics in each iteration, it is hoped that the metric-based method could generate solutions that are closer to optimal. The computational time would obviously increase with this iterative process. The development and testing of this iterative metric-based resource allocation strategy are presented in the following section.



(a)



(b)

Figure 5.4: (a) Strip-plot of the percentage difference between the objective function with solutions obtained from the numerical method and the metric-based method, grouped by m. The diamond markers are the averages of each group. (b) Scatter plot of the percentage difference between the objective function with solutions obtained from numerical method and the metric-based method verse the maximum value of $\Delta\mu_i$

5.2 Case Study 2 - Network with up to 6 Nodes

In this section, the proposed metric-based resource allocation strategy is applied to a collection of synthetic network models with 3 to 6 nodes. The optimisation problems for these synthetic networks are of higher dimensions than the optimisation problem in Case study 1. They can still be handled by the SLSQP function in SciPy but can no longer be solved analytically. The performance of the metric-based method and the numerical method are compared to see how well the metric-based method scale with the increasing dimensions.

As the simulations from the first case study suggest that the strategy works better with smaller portions of resources, this case study further developed the above-proposed ranking-based resource allocation to an iterative one, where the available resources are firstly divided into smaller portions before using the proposed metric-based method to distribute each portion.

5.2.1 The Network Models

To apply the proposed method, the network models need to have at least two layers, a service layer and an asset layer. The 141 connected graphs shown in Figure 5.5 are used as the asset layers, where each edge represents a railway line between two stations. The service layers are generated based on the asset layers.

For a given asset layer network from the 141 graphs, let n denote the number of nodes in the network. The maximum number of OD pairs the flow layer network can have is $n(n-1)/2$, which equals to the maximum possible number of node pairs in the asset layer network. The minimum number of OD pairs the flow layer should have is 1. Considering the asset layer networks in this case study are relatively small, the number of OD pairs, k, is set to have at least half of the maximum possible number of node pairs, and up to all of the possible node pairs. The set of k is all integers in this range, $k \in \mathcal{Z} \cap [\lfloor 0.5 * n(n-1)/2 \rfloor, n(n-1)/2]$. For each k, 10 separate service layers are generated by randomly selecting k number of node pairs from all of the node pairs. For each service layer, with its associated asset layer, the same method as in case study 1 is used to generate a weather parameter vector $\boldsymbol{\omega}$, a cost vector $\boldsymbol{c}$, a flow vector F, counting as a system configuration. 10500 system configurations are generated in total. For each system configuration, 4 different levels of resource input is considered, $m \in [0.25, 0.5, 0.75, 1]$, resulting in a total of 42,000 simulations.

5.2.2 The Numerical Method v.s. the Metric-based Method

For each system configuration, under each level of resource input m, the SLSQP minimisation function in SciPy and the metric-based method are used to find a solution each. The average percentage difference between the 42,000 pairs of solutions (10,500 systems configurations * 4 investment levels) is 6.8%, and the median is 5.3%. Figure 5.6a groups the percentage difference by the level of resource input, m, and plots them as a stacked histogram. The percentage

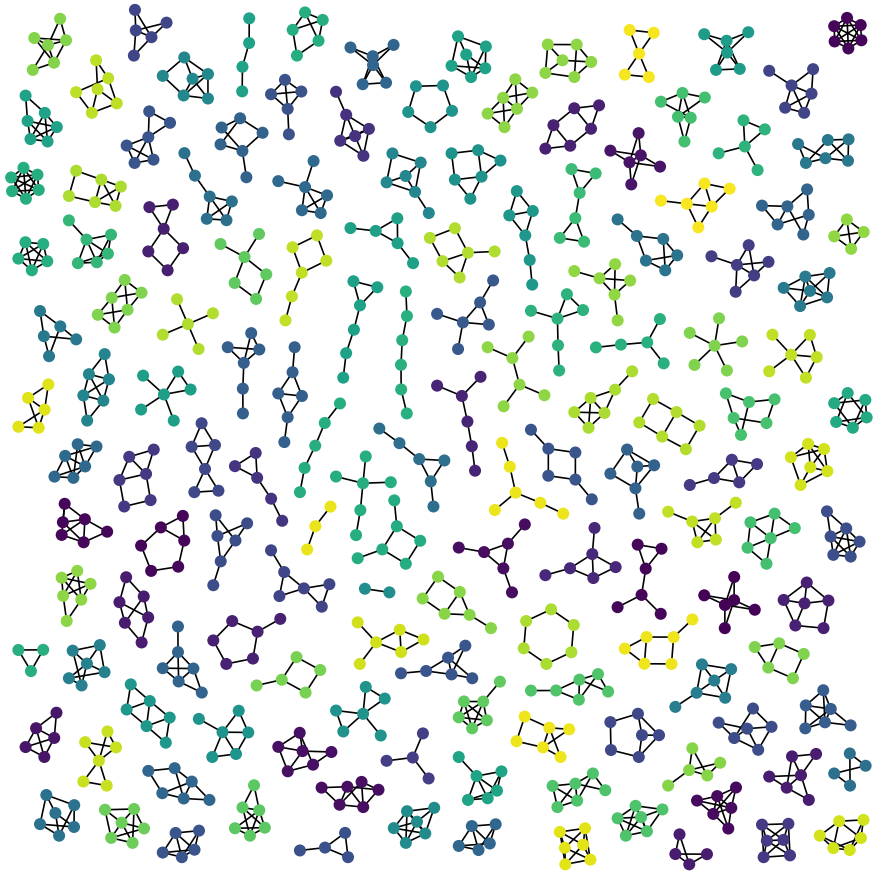
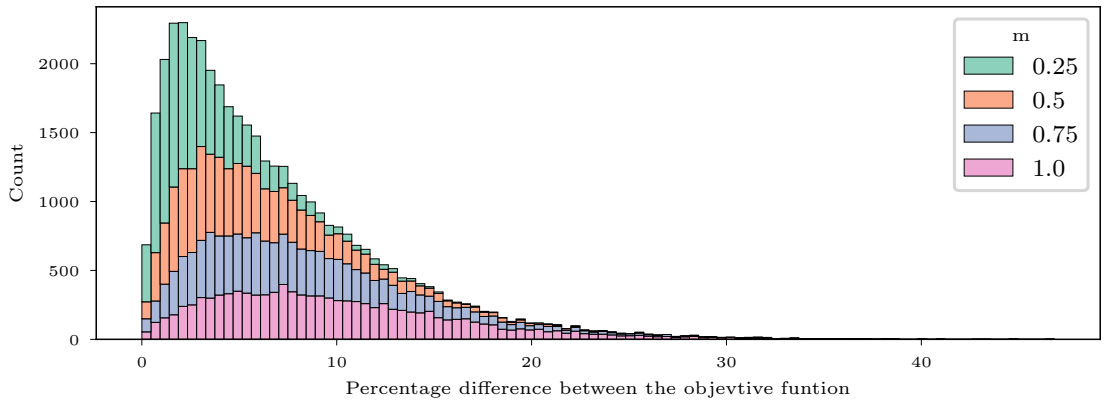


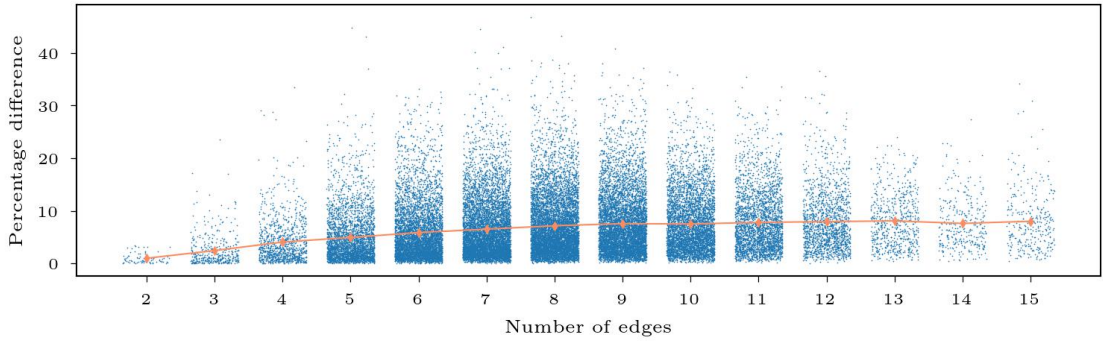
Figure 5.5: The graph atlas generated using codes from NetworkX (NetworkX, 2023)

differences between the objective functions overall increase with m, which is consistent with the finding from Case study 1.

Figure 5.6b shows the percentage difference between the objective function plotted against the number of edges in the asset layer, which is the number of dimensions of the optimisation problem. There are fewer asset layer networks with 10 or more edges than ones with less than 10 edges because only asset layer networks with 5 or 6 nodes can have more than 10 edges. Overall, the percentage difference increases with the number of edges in the network and has a tendency to plateau. This plateau means the percentage difference does not increase dramatically with the number of dimensions and the metric-based method has the potential to tackle high-dimensional problem with ease. It also means that when applying the method to high dimensional optimisation problems for large-scale infrastructure networks, the percentage difference between solutions generated with the metric-based method and the theoretical optimal solution can be estimated with a reduced-size network model.



(a) Stacked histogram



(b) Strip plot grouped by number of edges

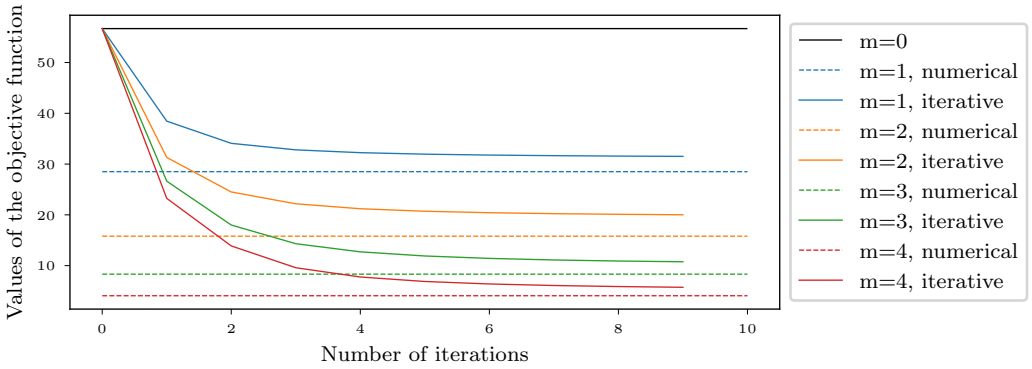
Figure 5.6: Percentage differences between numerical and metric-based method.

5.2.3 The Iterative Metric-based Resource Allocation Method

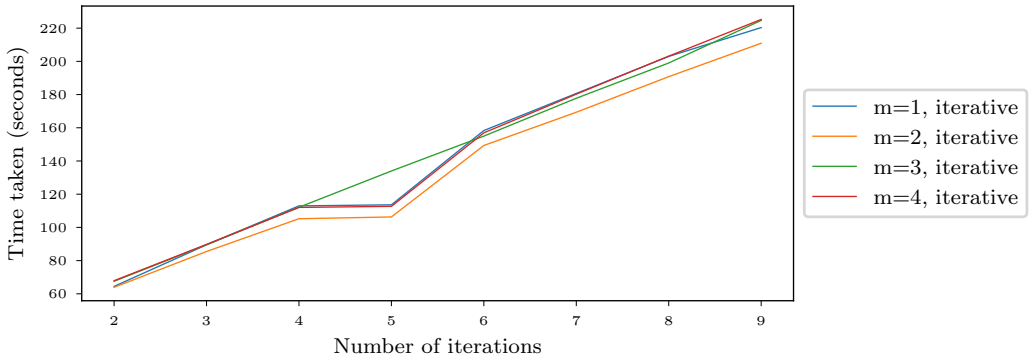
Case study 1 suggests that the metric-based method generates closer to optimal solutions with small amounts of input resources. An iterative metric-based method is hence developed and tested here. To implement the iterative process, the inputted resources need to be divided into smaller portions. There are two basic ways to achieve this: 1) specify the number of portions (e.g. divide the resources into 5 equal portions) or, 2) specify the size of each portion (e.g. each portion is of x amount of resources). Both methods were attempted in some preliminary simulation runs. This case study chooses to fix the number of iterations because the visualisation of results communicates better.

In the first iteration, the proposed metric is calculated with the system's initial configuration to allocate the first portion of upgrading resources. The solution from the first iteration is then added to the assets' current design load. In the second iteration, the proposed metric is calculated with the updated design load to allocate the second portion of resources. This process goes on until all iterations are complete.

The metric-based method naturally satisfies the constraints $\sum c_i \Delta\mu_i = C$ and $\Delta\mu_i \geq 0$. A check is added between the iterations to satisfy the $\Delta\mu_i \leq \Delta\mu_i^{ub}$ constraint. For any asset to which the allocated resources has reached the upper boundary, in all of the following iterations



(a)



(b)

Figure 5.7: a) Values of the objective function (percentage of service disruption) against the number of iterations used in the iterative metric-based method. The black line on top is the averaged initial service interruption without any upgrade (m=0). The dashed horizontal line are the objective functions minimised using SQP. The solid lines are the objective functions achieved by running the iterative metric-based method. b) Time taken (seconds).

its calculated metric will be reset to zero so that no more resource will be allocated to it. An upper boundary of $\Delta\mu_i^{ub} = 10$ is used in the absence of any relevant information, which can be revised when such information can be found.

Figure 5.7a plots the values of the minimised objective function against the number of iterations used for all four resource input levels (the solid curves). The objective function values are averaged across all 10,500 system configurations for each resource input level. The dashed horizontal lines are the objective functions obtained with the numerical solver SQP. It worth noting that the numerical method here can no longer guarantee that the solution found is the global minima (Boggs and Tolle, 1995). Due to the nonlinearity and nonconvex-nonconcave nature of the optimisation problem here, the solution found is for sure a local minimum, but it may not be the global minimum. However, it is the best possible solution the optimiser can find. From the plot, the difference between the numerical method and the iterative metric-based method decreases with increasing number of iterations. The rate at which the iterative metric-based method approaches the numerical solution decreases with increasing number of iterations. The iterative metric-based method can approach the numerical solution well.

In this method, the choice of the number of iterations matters. For a given amount of resources

to be allocated, a larger number of iterations will divide the resources into more and smaller portions. It is established that the metric-based method works better at allocating smaller amounts of resources as it is less likely to violate the assumption of a 'sufficiently small' increment. However, the simulation results suggest that increasing the number of iterations can only be beneficial to a certain degree because the time taken increases linearly with the number of iterations as shown in Figure 5.7b. An overly large number of iterations is not the best choice. For a given problem, there may exist an optimal number of iterations to use, which can be found using techniques such as elbow analysis.

5.2.4 Discussion

This section applies the proposed metric-based resource allocation strategy to a collection of synthetic network models. The differences between the objective functions of the proposed method and the optimiser tends to plateau with the increasing number of dimensions. The proposed method scale well with the increasing dimensions and can still generate near-optimal solutions even when facing high-dimensional problems. With an increasing number of iterations, the gap between the iterative metric-based method and the numerical solution gets smaller. However, it would need a well-chosen number of iterations to balance the solution optimum and computational time. Next section applies the developed iterative metic-based method to a rail freight network.

5.3 Case Study 3 - UK Core Freight Rail Network

Rail freight plays a key role in the UK economy. In the past five years, annual average freight moved by rail was 16.6 billion net tonne kilometres (ORR, 2023), delivering £2.45bn of benefits to the UK's economy (Deloitte, 2021). This section presents the application of the proposed resource allocation approach to the UK's core freight network to illustrate the applicability of the iterative metric-based resource allocation method to large-scale infrastructure systems.

5.3.1 The Objective Function

To recap, the optimisation problem is to find the vector $\boldsymbol{\Delta\mu}$ that minimises the expected onset service disruption:

$$\min_{\boldsymbol{\Delta\mu}} \sum_j F_j \left(1 - \prod_i \left(1 - H_{i,j} \cdot \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{\omega_i - (\mu_i + \Delta\mu_i)}{\sigma_i\sqrt{2}}\right)\right)\right)\right), \tag{5.31}$$

subject to:

$$\sum_{i=1}^{|\mathcal{E}_\alpha|} c_i \Delta\mu_i = C, \tag{5.32}$$

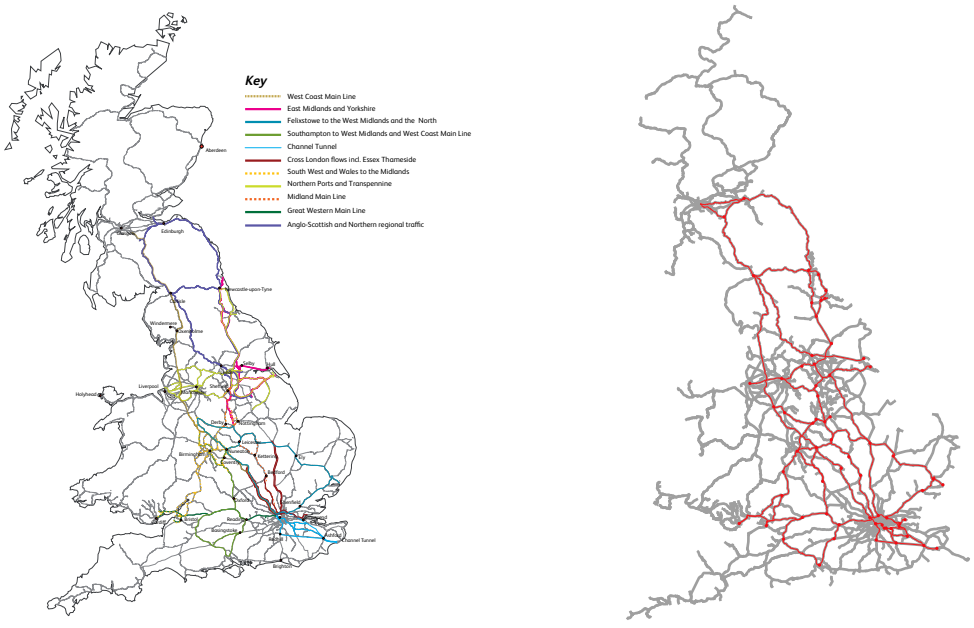
$$\Delta\mu_i \geq 0, \quad i = 1, 2, \cdots |\mathcal{E}_\alpha|, \tag{5.33}$$

$$\Delta\mu_i \leq \Delta\mu_i^{ub}, \quad i = 1, 2, \cdots |\mathcal{E}_\alpha|, \tag{5.34}$$

where F is the flow vector, H is the route-link incidence matrix, $\boldsymbol{\omega}$ is the weather condition vector, $\boldsymbol{\mu}$ is the expected design load, $\boldsymbol{\sigma}$ is the variance of the design load, $\boldsymbol{c}$ is the cost vector, and $\Delta\boldsymbol{\mu}$ is the solution vector. From here in this chapter, the values of expected onset service disruption as a percentage of the normal amount of flow will be referred as *the objective function*.

5.3.2 The Network Model

The rail freight system network is shaped by the market's demand for transporting certain commodities from one location to another. These demands drive the formation of some critical corridors for the movement of those commodities (Network Rail, 2017a). These critical corridors collectively form a core freight network. This case study focuses on a core rail freight network comprised of nine freight corridors (Figure 5.8a) identified in a Freight Network Study (Network Rail, 2017a).



(a) Key freight corridors by Network Rail (b) The freight network in this study

Figure 5.8: (a): The key freight corridors identified by Network Rail Network Rail (2017b). (b): The asset layer constructed for this study (red) embedded in all railway track lines across the country that is managed by Network Rail (grey). The original shapefile is provided by Network Rail.

A system model for this core freight network is constructed, consisting of two layers - an asset layer and a service layer. The edges in the asset layer represent the railway track lines. With the

aid of the maps and descriptions of the freight corridors in the Freight Network Study (Network Rail, 2017a), the asset layer is constructed by manually selecting track lines and stations from the shapefile that contains all track lines across the country (Figure 5.8b). The asset layer consists of 65 nodes and 84 edges.

The service layer is constructed using the freight train schedule data obtained from the open data feed (openraildata.com) by Network Rail (downloaded on March 2022). The freight train schedule data includes train schedules created as part of the Long Term Planning (LTP) and Short Term Planning (STP) processes. Only the 'permanent' schedules, the base schedule created as part of the LTP process, are used to construct the service layer. The short-term variations and planned cancellations are omitted as this study is focused on the resilience of the network in the long-term.

The next step is to identify the *core* freight train schedules out of all schedules, which is done by identifying freight train schedules that run between stations in the asset layer. In the obtained train schedule data, each schedule has a unique identifier and details all the stations the train originates at, passes, stops at and terminates at. This list of stations is anointed as 'location record'. The location record is firstly reduced to stations that the train originates at, stops and terminates at, by removing stations that the train passes without stopping.

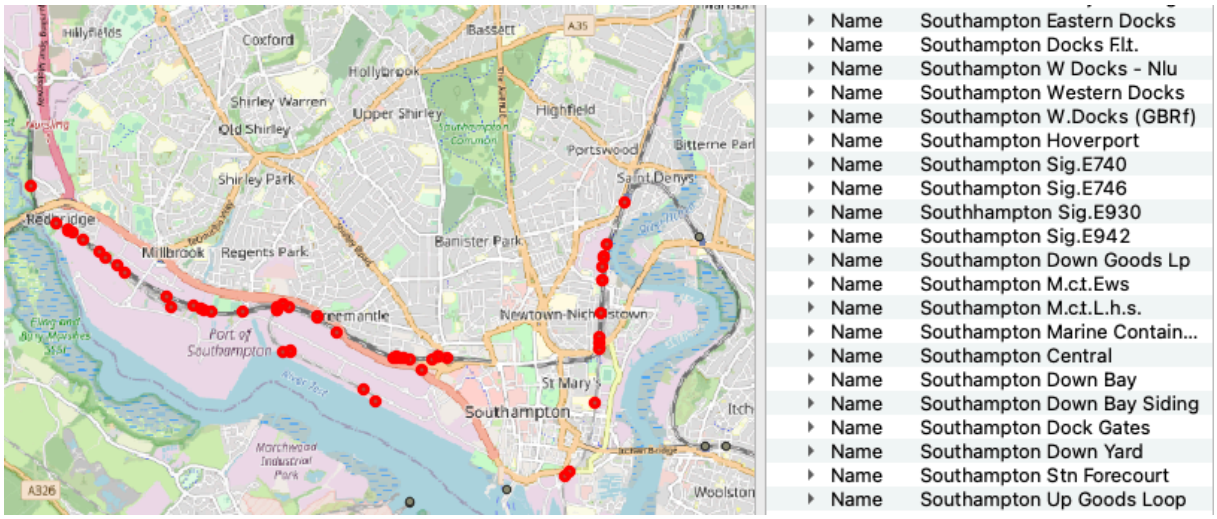


Figure 5.9: Cluster of stations near Southampton

The location record is further simplified and shortened to contain only stations in the core freight network (the asset layer). One of the challenges encountered is that a station in the constructed core freight network represents a group of stations in the train schedule data. For example, as shown in Figure 5.9, the 'Southampton' station in the core freight network in fact represents a cluster of stations in the train schedule data. To simplify the location record, for all stations in the location record, the station name is replaced with the name of the closest station in the asset layer if the distance between them is less than 5km. Otherwise, it would be left empty. If a schedule's location record is simplified to an empty list or a list of only one station, it means this schedule is not running on the track lines of the asset layer. The minimum requirement

for a schedule to be considered as a core freight service is that it originates and terminates at core stations. Such simplification leads to some schedules having the same list of stations, which will be merged and regarded as one OD pair. The constructed service layer has 1623 OD pairs. Converting the simplified location records to node paths in the asset layer, then to edge paths, gives the link-route incidence matrix, H.

The schedule data also contain the start and end date of the planned schedule and a field called 'schedule days runs' indicating which days of a week the schedule runs. The field has seven characters, where the first character denotes Monday and the seventh denotes Sunday. If a character on a position is the number 1, it indicates that the service operates on that day, whereas a 0 means it does not. In practice, each freight train schedule has multiple calling points between the origin and destination station for loading and unloading goods. Due to a lack of information on the amounts of goods loaded and unloaded at each station, the 'goal' of the service layer in this model is simplified to 'complete the journey from the origin node to the destination node'. As the downloaded dataset does not include information on the weight or volume of goods transported, a scheduled run on a single day is treated as 1 unit of flow in the service layer. The total number of runs for each schedule from April to September is firstly counted. The total number of runs is then divided by the number of days from April to September to give the average summer daily traffic profile. Adding up the trip of all schedules whose simplified location records are identical gives F, the amount of flow between the OD pairs.

With all efforts that can be possibly made, no asset information is available to support the construction of the cost vector, $\boldsymbol{c}$. The cost vector is hence assumed to be proportional to the geographical length of the edges. The underlying assumption is that all track lines are of the same asset condition, and the unit upgrading cost for an edge is directly proportional to the length of the railway lines it represents.

5.3.3 The Climate Hazard and Climate Data

This case study uses the high temperature-related track buckling as the example hazard. The same fragility function as in Chapter 3 Section 3.3.1 is used, which states:

$$p_i = \xi(\omega_i) = \frac{1}{2}\left(1 + \text{erf}\left(\frac{\omega_i - \mu}{\sigma\sqrt{2}}\right)\right) \tag{5.35}$$

where p_i and ω_i is the probability of failure and local near-surface air temperature for asset i; ω is the critical air temperature that the track was designed to withstand, and μ is associated with the variation between the actual critical air temperature the track can stand and the critical air temperature it was designed to withstand. The value of ω and μ for all of the edges are set to 35°and 2.5 in this case study, under the assumption that all tracks in the network are of the same condition. This assumption was made without other choices as no information on asset conditions was available by the time this work was done. The values of ω and μ can be updated to asset-specific ω_i and μ_i in future applications when possible.

The next step is to select future failure scenarios to optimise toward. Two approaches are used to select two collections of daily profiles from the summer days (May to September) of the years 2051-2100. The first collection of days contains the 5 hottest days of each year and therefore 250 days in total for the 50 years. These extreme days are times when the railway system is under the highest level of pressure and is likely to experience some of the worst disruptions. Under such situations, a good resource allocation could differentiate itself from a mediocre one.

The second collection of days is a collection of typical days instead of extreme days. It is selected using a clustering method where the summer days for 2051-2100 are clustered together to select 250 representative example days, 250 is used to keep the same number of days as the collection of extreme days. The 153 summer days from 1st May to 30th September for the 50 years, which gives 7650 days in total, are firstly clustered together using a bisecting KMean algorithm. After clustering the summer days into 15 clusters, 251 days are selected by randomly choosing x number of days from each cluster, where x is the nearest integer of the size of the cluster multiplied by 250 divided by 7650. The typical day collection end in 251 days.

5.3.4 Numerical Solve: Heuristic Algorithm

In searching for a computational intelligence-based algorithm that can be effectively applied to power system optimisations, Del Valle et al. (2008) finds the PSO algorithm the most suitable choice for several reasons. 1) The PSO is not largely affected by the non-linearity of the problem and scales well with complexity. 2) The implementation is straightforward. 3) compared to other heuristic algorithms, such as the generic algorithm, PSO has fewer parameters to adjust. 4) All particles remember their previous best and use the population best for their next move. Considering the optimisation problem in this work is non-linear, continuous, high-dimensional, and with both equality and non-equality constraints, for the same reasons are in Del Valle et al. (2008), the Particle Swarm Optimisation (PSO) algorithm (Kennedy and Eberhart, 1995) is a natural choice. A Python implementation of the PSO (pyswarm, 2015) is used to solve the optimisation problem here.

5.3.5 The Iterative Metric-based Solve

The preliminary results in Section 5.2 show that the proposed metric-based method can minimise the objective function very close to the numerical method by adopting an iterative resource allocation method. The idea is to firstly divide the available resources into multiple smaller portions and run the metric-based approach repeatedly to allocate each portion. As discussed in Case Study 2, when implementing an iterative process, there are two ways to divide the resources - 1) specify the number of iterations to run, and 2) specify the size of each portion, the stepsize.

In Case Study 2, it was chosen to fix the number of steps and then calculate the amount of resources to be allocated in each iteration. To satisfy the upper boundary constraint, in

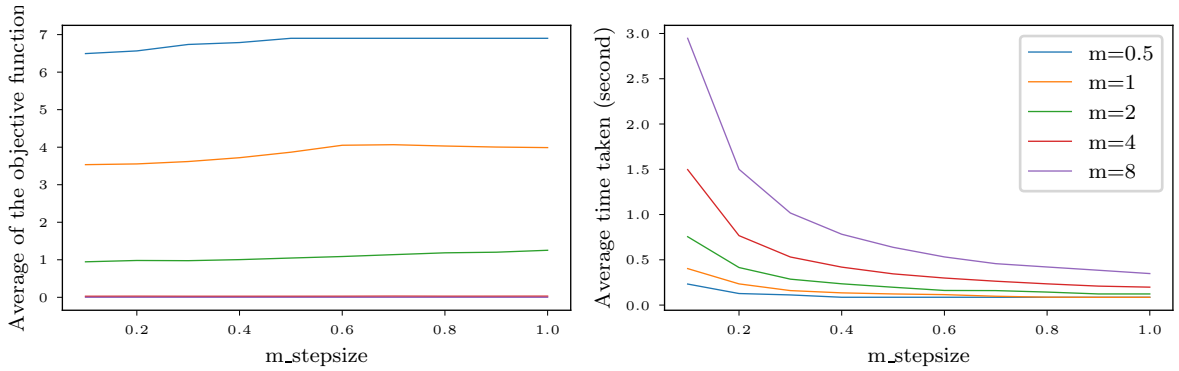
each iteration, if the allocated resource to an edge violates the upper boundary constraint $\Delta\mu_i > \Delta\mu_i^{ub}$, the amount of resources allocated is then adjusted to $\Delta\mu_i = \Delta\mu_i^{ub}$ to satisfy the upper boundary constraint. However, this simple adjustment in fact leaves a certain amount of resources wasted. This is because when it resets the values of $\Delta\mu_i$ to $\Delta\mu_i^{ub}$, there is $c_i * (\Delta\mu_i - \Delta\mu_i^{ub})$ amount of resources wasted in the current iteration and is not recycled to be used in any following iterations. This leads to a small fraction of resources unused in the case study in Section 5.2 as the networks are small.

When directly adopting this method to the freight network, it is noticed that there are cases where the unused resources are as high as 80%. Therefore, some improvements have been made. First, instead of specifying the number of iterations, a stepsize *m_step* is specified. In each iteration, the algorithm tries to allocate *m_step* amount of resource to the edges. The solution obtained will be checked, and where there is a violation of the upper boundary constraint, the value of $\Delta\mu_i$ will be reset to $\Delta\mu_i^{ub}$. The unused resources $c_i * (\Delta\mu_i - \Delta\mu_i^{ub})$, is recycled back to the reservoir of not-yet-allocated resources and will be allocated in the following iterations. The iteration goes on and in each iteration it will attempt to allocate *m_step* amount of resources as long as the reservoir still has more than *m_step* amount of resources. If there is less than *m_step* amount of resources in the reservoir, the algorithm will allocate the residual until all resources are used. The pseudo-code explaining the iterative metric-based resource allocation is presented in Algorithm 2.

ALGORITHM 2 - The Iterative Metric-based Approach

- INPUT
  - m - the total amount of resource
  - m_step - stepsize for m,
  - ω - the weather parameter vector
  - μ - the current design load vector
  - H - the link-route incidence matrix
  - c - the unit cost vector
  - ub - upper boundary for the solutions
- INITIATE
  - $\Delta\mu = 0$
  - $utilisation\_ratio = \sum(\Delta\mu * c)/m/\sum c$
  - $reach\_ub = [\ ]$
- WHILE $utilisation\_ratio < 0.999$:
  - $unused\_resource = m * (1 - utilisation\_ratio)$
  - $mi = min(m\_step, unused\_resource)$, resource to be allocated in this round
  - calculate the metric, χ
  - $\chi[i] = 0$ if $i \in reach\_ub$, (exclude edges who reaches their ub)
  - $x = mi * \sum c * (\chi/\sum \chi)/c$
  - $reach\_ub.append(i)$ if $x[i] >= ub[i] - \Delta\mu[i]$
  - $x[i] = ub[i] - \Delta\mu[i]$ if $x[i] > ub[i] - \Delta\mu[i]$
  - $\Delta\mu \ += x$
  - $utilisation\_ratio = \sum(\Delta\mu * c)/m/\sum c$
- RETURN $\Delta\mu$

To find the optimal value of stepsize, *m_stepsize* for this case study, the 5 extreme days for the year 2051-2055 are used as testing days. 25 days are considered in total. Five investment levels are considered, $m \in [0.5, 1, 2, 4, 8]$ and 10 stepsize are tested, $m\_step \in [0.1, 0.2, \cdots, 0.9, 1.0]$. For each day, under each investment level, the iterative metric-based approach allocates a specified amount of resources using each of the stepsize. Figure 5.10 shows the relationship between the values of the objective function obtained and the time taken with different values of *m_step*. Though the changes in objective function value is not very obvious with *m_step*, there is a region of linear change on the left half of the plot in Figure 5.10 (a). The time taken increases non-linearly with smaller values of *m_step*. After carrying out elbow analyses to find the 'turning points' of the curves on Figure 5.10 (b), $m\_step = 0.2$ is used for this case study. For any given amount of resource input, in every iteration, the iterative process will allocate $m = 0.2$ amount of resource (or less when there is less than $m = 0.2$ amount of resources left) until all inputted resources are allocated.



(a) (b)

Figure 5.10: (a) The average of the objective functions from all 25 days versus the values of *m_stepsize*. b): The average of the simulation times for all 25 days versus thee values of the *m_stepsize*

5.3.6 Results - Heuristic v.s. Iterative Metric-based

Using the 5 hottest days each year for the years between 2051 and 2100, the performance of the PSO algorithm and the iterative metric-based method is compared in terms of the time taken and the values of the objective function. Five days each year for a 50-year period gives 250 weather profiles. For each weather profile, 5 investment scenarios are considered, $m \in [0.5, 1, 2, 4, 8]$, which gives 1250 cases to solve. To solve the 1250 cases, it took the PSO algorithm 18291 seconds (∼5 hours) to complete without parallelisation with 8GB RAM. It took the iterative metric-based method 855 seconds (∼15 minutes) to solve without parallelisation with 8GB RAM. The iterative metric-based method solves the optimisation problem significantly faster than the PSO.

Figure 5.11 shows the objective function solved with both methods against the amount of resource input. The objective function is the expected initial service disruption in the system when subjected to a given weather profile as a percentage of the service flow amount with no

asset failures. The method concludes lower objective function value is considered optimum. From Figure 5.11, except $m = 0$ with zero investment and $m = 8$ with excessive investment, the proposed iterative metric-based method consistently gives an objective function that is smaller than the PSO algorithm when given the same amount of resources to distribute between assets.

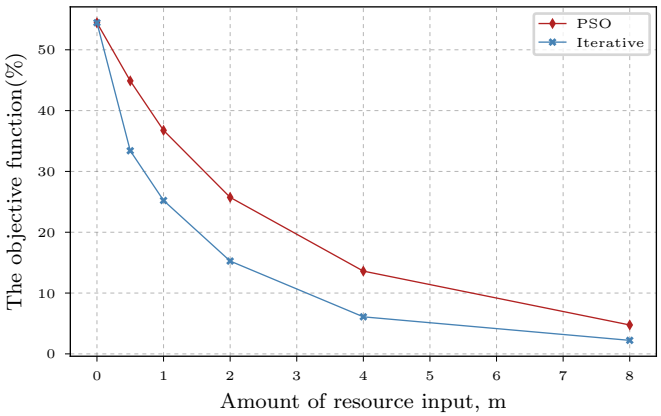


Figure 5.11: The average of the objective function minimised using the PSO algorithm and the proposed metric-based approach with different levels of investment ($m \in [0.5, 1, 2, 4, 8]$). The objective function is the expected initial service disruption in percentage of the normal service level for a given weather profile. m indicates the investment level. $m = 2$ means the amount of investment equals to that required to increase the design load for all assets in the network by 2°C.

A degree of correlation is expected from two methods as they are optimising toward the same goal. To evaluate the correspondence between the resource allocation solutions found using the two methods, the Kendall rank correlation method (Kendall, 1970) is used to measure the rank correlations between the solutions. It is one of the most widely used rank correlation statistics (Puth et al., 2015), which compares the ranking of elements of two data sets. To carry out this statistical testing, the solutions obtained need to be converted to a rank format. The direct way for this conversion is to rank the edges by $\Delta\mu_i$ in the solution vector and compare the ranks from two solutions. However, the solutions in this case study are continuous variables, and the rank based on the actual values could be oversensitive. Further, the precision used in relevant documents reviewed rarely uses any decimal places. Therefore, the $\Delta\mu_i$ are rounded to the nearest integer before being ranked.

In the converted rank, two or more edges could have the same rank value (called *tie*). Therefore, the Kendall's τ-b coefficient, which makes adjustments for ties, is used here. The value of the Kendall's τ-b coefficient range from -1 (strong disagreement, 100% negative association) to 1 (strong agreement, 100% positive association.) From Figure 5.12, there is a positive association between the resource allocation solutions produced by the PSO algorithm and the proposed metric-based method. The positive association suggests an agreement between the rankings when fewer resources are available. This is in line with the main assumptions made by linearizing the fragility function around the failure load. Also, it is far more consistent with the reality of the potential implementation, when usually only limited resources are available to allocate. Clearly, when a large amount of resources become available, the range of possibilities expands and multiple allocation strategies could become equally optimal, which leads to the generally

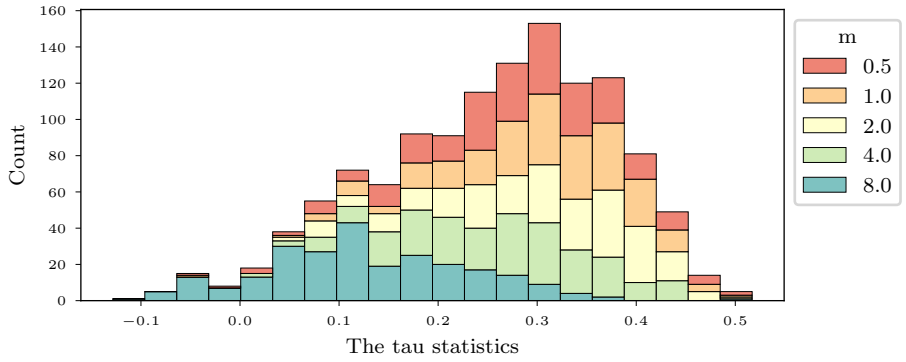


Figure 5.12: Histogram showing the calculated Kendall's τ-b coefficient between the solutions found by the PSO algorithm and the iterative metric-based method.

uncorrelated ranking for very high values of m. As shown later in Figure 5.14, when m increases to 8, more edges reach the upper boundaries using the metric-based method. Due to the heuristic nature of the PSO algorithm, it's less likely to always find solutions precisely on the upper boundary. For instance, it's more likely to be a value very close to 10, like 9.998, but not exactly 10. While the difference between 10 and 9.998 may seem small, it can result in discrepancies in the tau-statistics.

5.3.7 Results - Optimising towards Extreme Days and Toward Typical Days

The iterative metric-based method was applied to two distinct collections of days. The first collection of days comprises the 5 hottest days each year for the year 2051-2100, which will be referred to as *the extreme days*. The second collection of days comprises 251 days selected using a clustering method as described in Section 5.3.3, which will be referred to as *the typical days*. For each day in a collection, a solution is generated using the iterative approach. The solutions generated for all days in a collection of days are averaged along each edge to give a combined solution.

$$\begin{array}{c} \\ e_1 \\ e_2 \\ \vdots \\ e_n \end{array} \begin{array}{c} \boldsymbol{x_1} \\ \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1n} \end{bmatrix} \end{array} + \begin{array}{c} \boldsymbol{x_2} \\ \begin{bmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2n} \end{bmatrix} \end{array} + \cdots + \begin{array}{c} \boldsymbol{x_n} \\ \begin{bmatrix} x_{n1} \\ x_{n2} \\ \vdots \\ x_{nn} \end{bmatrix} \end{array} = \begin{array}{c} \boldsymbol{x_{comb}} \\ \begin{bmatrix} (x_{11} + x_{21} + \cdots + x_{n1})/n \\ (x_{12} + x_{22} + \cdots + x_{n2})/n \\ \vdots \\ (x_{1n} + x_{2n} + \cdots + x_{nn})/n \end{bmatrix} \end{array}$$

Figure 5.13: Combine the solutions by averaging along each edge.

Figure 5.14 (a) and (b) show the combined solutions with bar plots, where each bar represents an edge and is positioned by the latitude of the centroid of the edge. From $m = 0.5$ to $m = 8$, with an increasing level of resource input, the amount of increase in design load between edges gradually levels up resulting in a decrease in the inequality in resource allocation. To quantify the inequalities between the edges, the Gini coefficient (Gini, 1912) is calculated for the averaged solutions under each investment scenario. It is the most widely used measurement for household

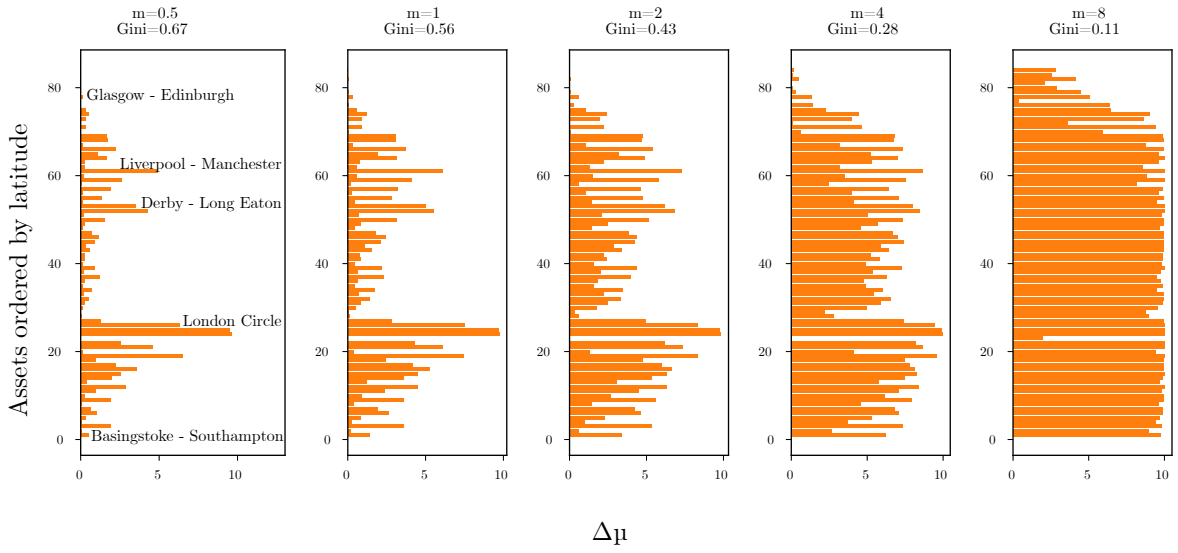
income inequality (ONS, 2022) and is applicable for measuring the resource allocation inequality in this case study.

In this case study, the Gini coefficient measures the inequality in $\boldsymbol{\Delta\mu}$ and takes a value between 0 and 1. A Gini coefficient of 0 indicates that the values of $\Delta\mu_i$ in $\boldsymbol{\Delta\mu}$ are perfectly equal. A Gini coefficient of 1 indicates that all the available resource is allocated to one single edge, leaving the rest of the network without any upgrade. The optimised solution toward the extreme days have a Gini coefficient of 0.67 at an investment level of $m = 0.5$. As the level of investment increases, the inequality decreases and reaches a Gini coefficient of 0.11 at $m = 8$. A similar trend is shared by the optimised solutions toward the typical days. It implies that when resources are limited, concentrating them on fewer but more critical assets could be more beneficial rather than distributing the resources more widely.

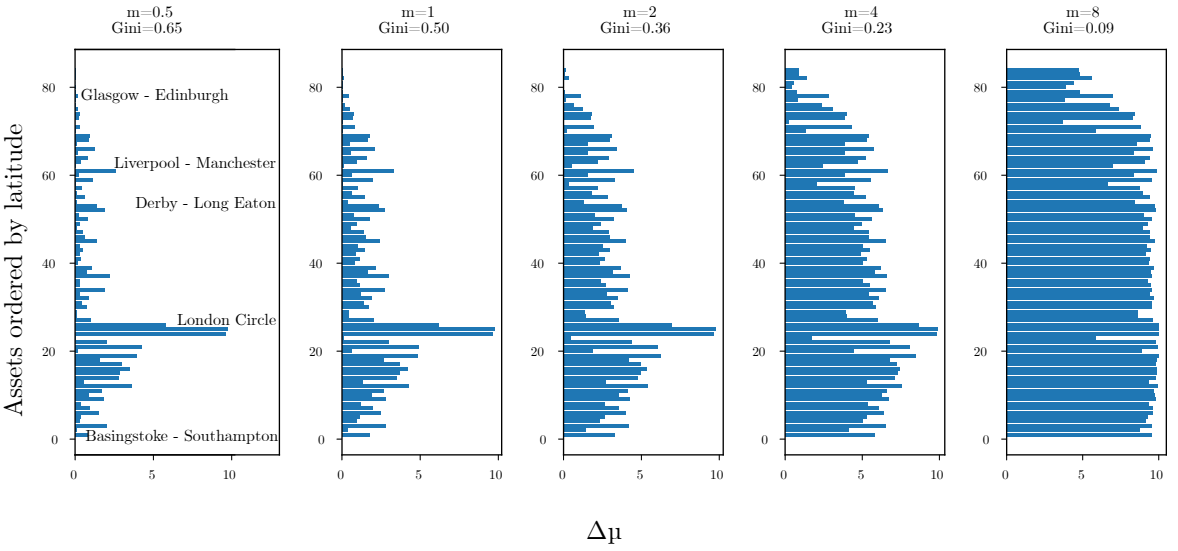
Looking only at the bar plots for $m = 0.5, 1$ and 2, in Figure 5.14 there are two peaks in the bar plots showing the combined solution from optimising toward the extreme days. Whilst there is only one peak in the bar plots for the typical days. When subjected to the same level of investment, the optimised solutions toward the extreme days consistently have a slightly higher Gini coefficient than the optimised solutions toward the typical days. This difference could be explained with the finding that limited resources is likely to result in high inequality in asset upgrading. The typical days set comprises a range of weather patterns, for example, days where the temperature is below the critical temperature everywhere with near-zero disruption, days where the south regions are significantly hotter than the north, and days where it is hot across the country. Whilst the collection of extreme days only comprises the hot days. The level of resource scarcity is relative to both the severity of the event (by how much the external weather parameter exceeds the current design load) and the amount of upgrade resources available. The same amount of resource input is more likely to be considered 'limited' for a day in the collection of extreme days than a day in the collection of typical days.

5.3.8 Results - Correlations with Constant Variables

The resource allocation process here is not a complete black box process as the relationships are known and expressed as functions. In each iteration, the solution correlates to the proposed metric when not considering the upper boundary constraint. With the enforcement of the constraints and the adding-up of solutions from each iteration, it is worth checking the correlations of the final solutions with some known variables. Figure 5.15-5.18 show the correlations between the solutions and three known input constants - the latitude of the edges, the reciprocal of the geographical length of the edges (the cost vector), and the flow throughput. The scatter plots reveals the shapes of the correlations, which varies from case to case and makes it difficult to compare the strength of correlations with a single type of correlation analysis. The Kendall's τ-b coefficient is used to compare the ranking of edges in the solutions and the ranking of edges by those known constants. The solutions are rounded to the nearest integer as discussed in Section 5.3.6.



(a) The combined solution from optimising toward the extreme days



(b) The combined solution from optimising towards the typical days

Figure 5.14: Bar plots showing the combined resource allocation solutions in relation to their latitude. Each bar represents an edge in the network, and the length of the horizontal bar represents the amount of increase in its design load. The bar plot organises the edges by the latitude of their centroid. The south-most edge is at the bottom of the plot. The y-axis is asset's rank by latitude and is not actually its latitude.

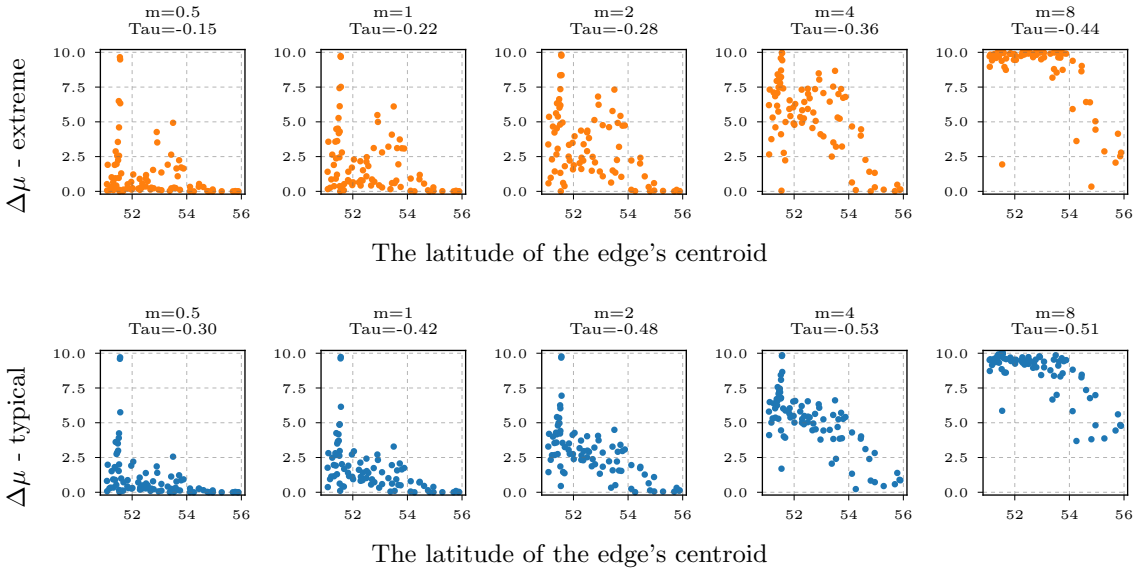


Figure 5.15: Correlation between the solution and the latitude of the edges.

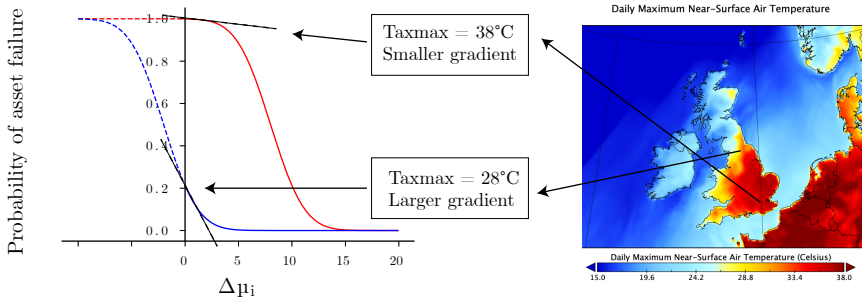


Figure 5.16: The decrease of probability of failure against increasing $\Delta\mu_i$ for when the local tasmax is 38°C (red) and 28°C (blue).

From Figure 5.15, there is a weak negative correlation between the solutions and the latitude of the edges, for both the case of optimising toward the extreme days (orange) and the typical days (blue). The latitude of edges can be interpreted as a rough indicator of temperature and hence the probability of failure. Edges in the south are of lower latitudes and is more likely to experience high temperatures than the north. The strength of this correlation is not very strong when the resource input level is low, meaning the climate-related failure probability is not the most significant factor among the three input constants considered. However, the strength of this correlation increases with the amount of resource input, meaning the climate-related failure probability starts to play a more important role in directing resource allocation. Originally a strong correlation was expected between the amount of upgrade and the latitude of the edges as it is generally accepted that the south would suffer more than the north with global warming. One possible explanation for the observed weak correlation could be that the probability of failure for southern assets are so high that the benefit of investing in those edges is diminutive compared to investing in edges in the north. Whilst in the north, conditions are less severe

meaning investing in edges is more likely to result in some notable increase in the expected delivery. Figure 5.16 plots the decrease of probability of failure against increasing $\Delta\mu_i$ for when the local tasmax is 38°C (red) and 28°C (blue). Investing in the high-temperature zone edges could be less beneficial than investing edges in the lower-temperature zone.

From Figure 5.17, some strong positive correlation can be observed and supported by the calculated Kendall's τ-b coefficients. The strong correlation can be easily interpreted by shorter edges being obviously cheaper to upgrade. In the absence of detailed asset information, the geographical length of the edges were used as the cost vector by assuming that all track lines have the same asset condition and the cost to upgrade an edge is proportional to the length of the railway lines it represents. From Equation 4.121, the unmodified solution in any single iteration should be proportional to the reciprocal of the edge's unit cost.

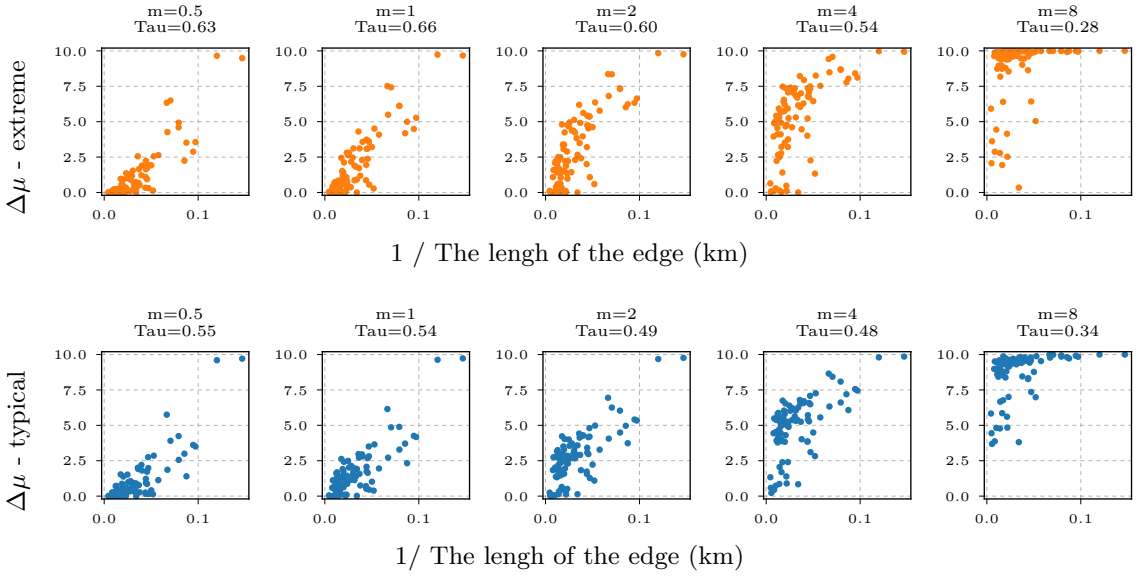


Figure 5.17: Correlation between the solution and the unit cost of the edges.

Figure 5.18 shows the correlations between the solutions and the flow throughput. From the scatter plots and the calculated Kendall's τ-b coefficients, there is a moderate level of correlations between the solutions and the flow throughputs. The flow throughput is the amount of flows, or services that passes through the edge,which in fact is the OD centrality metric. The flow throughput of an edge represents the amount of flow drop if the edge fails and to an extent indicates the liability of the edge.

Overall, the solutions found correlates to the latitude the least, with the absolute values of the Kendall's τ-b coefficients averages at 0.37. The solutions found correlate to the reciprocal of the unit cost and the flow throughput equally with the Kendall's τ-b coefficients averages at 0.51 and 0.52. This finding is reflected in the map visualisation of the optimisation solutions. Figure 5.19 and Figure 5.20 maps the optimised solution to the geographical shapefile of the edges in the core freight network model. For both the extreme days and typical days maps, when resources are limited, e.g. $m = 0.5$ and $m = 1$, the unexpected hotspot up in the north

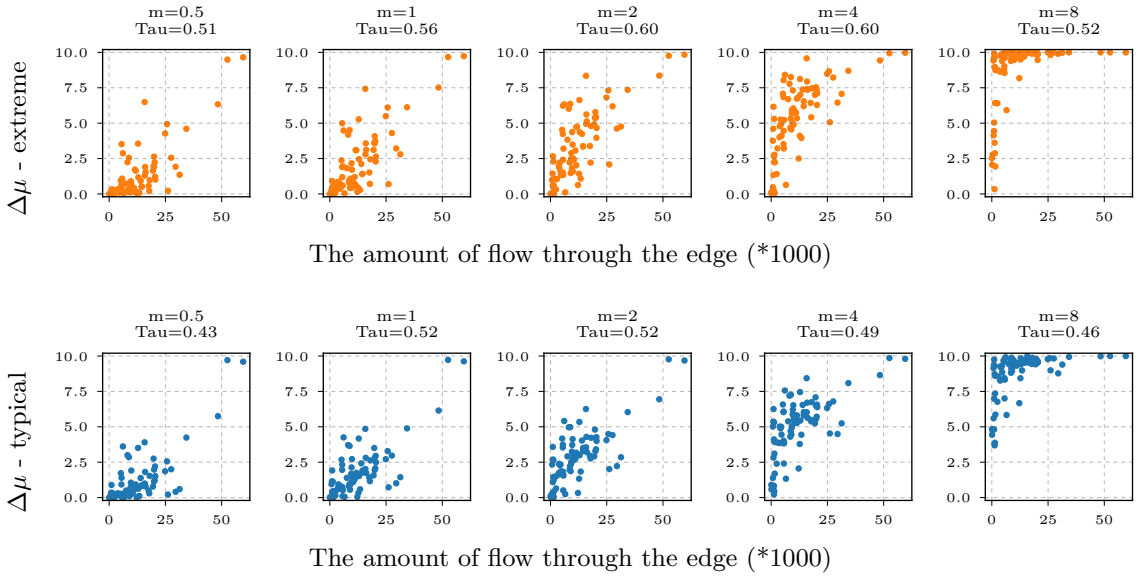


Figure 5.18: Correlation between the solution and the flow throughputs of the edges.

suggests that the unit cost and/or the flow throughput of edges could outweigh the probability of failures of assets in the resource allocation.

If to compare the extreme days maps with the typical days maps, greater differences among edges' line thickness can be noticed in the maps of extremes days, pointing to higher inequality level. This is consistent with the calculated Gini coefficient in Figure 5.14 in Section 5.3.7. As discussed before, the difference in equality could be explained by the different constitution of daily weather patterns between the two set of days and the fact that resource scarcity leads to inequality.

5.3.9 Discussion

This case study applies the iterative metric-based method and a PSO algorithm to a core railway freight network of GB. The simulation suggests that the proposed iterative metric-based method can conclude better solutions in shorter computational time. The proposed method in essence is a cost/benefit analysis. The case studies in a way shows that the cost/benefit analysis, if carried out in a systematic way, has the potential to match, or even outperform some of the computationally expensive optimisation algorithms.

As both methods are optimising toward a shared objective function, initial expectations were that the solutions would demonstrate a reasonable degree of correspondence. However, the correlation quantified using Kendall's τ coefficient is not very strong, which is attributed to the underperforming of the PSO. The reasons including 1) The PSO is free to search the whole solution space, which contains the optimal solution; 2) in case study 2 when the optimisation problems are of lower dimensions, which can still be handled by the SciPy package, the numerical method performs better than the metric-based method. The objective function minimised

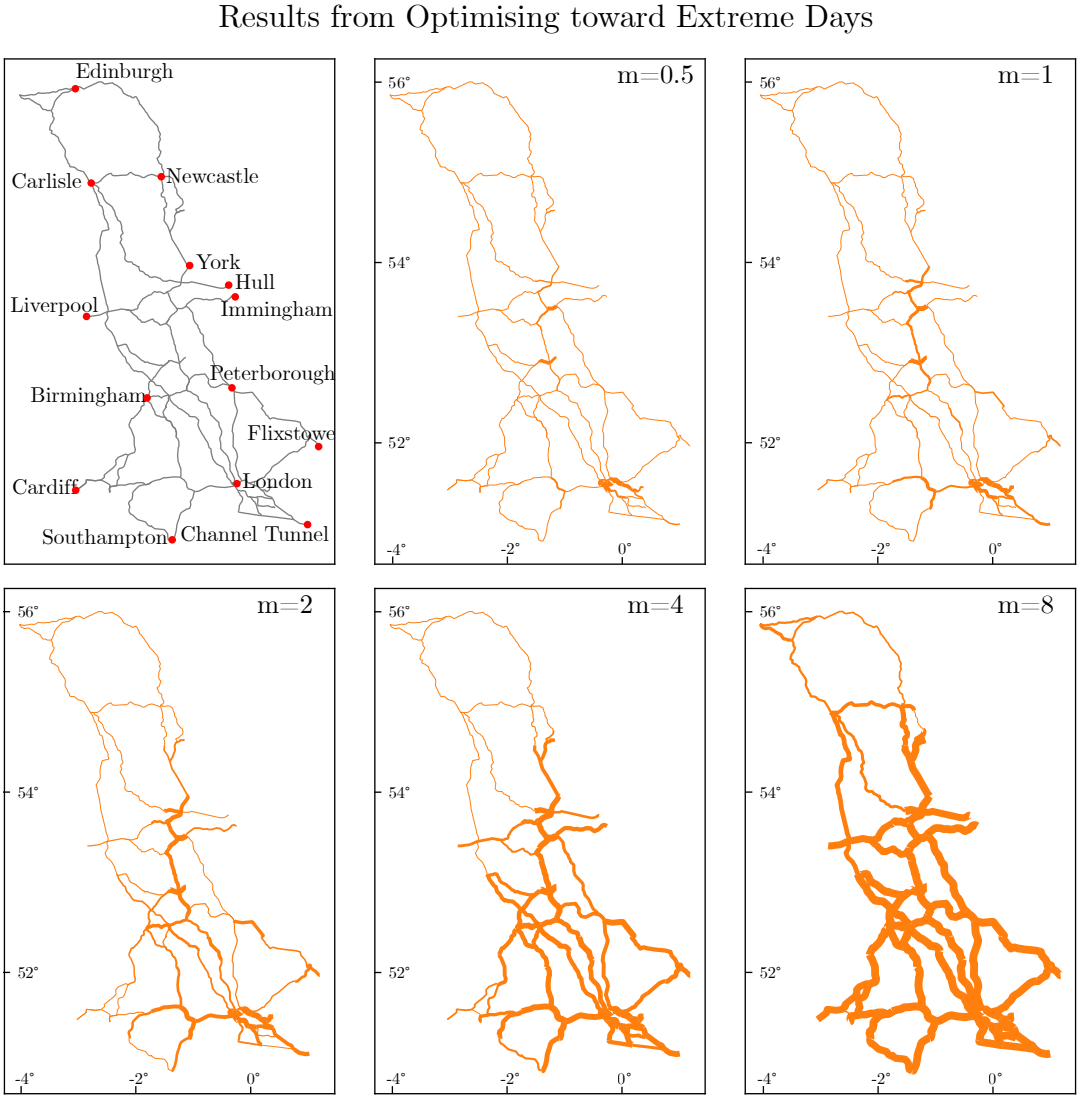


Figure 5.19: Maps showing the resource allocation solutions obtained from optimising toward the 5 hottest day each year for the years 2051-2100. Width of the lines indicates the increment in design load $\Delta\mu_i$ for the asset. m is the variable used to quantify the amount of investment. $m = 2$ means the amount of investment equals the amount required to increase the design load for all assets in the network by 2°C.

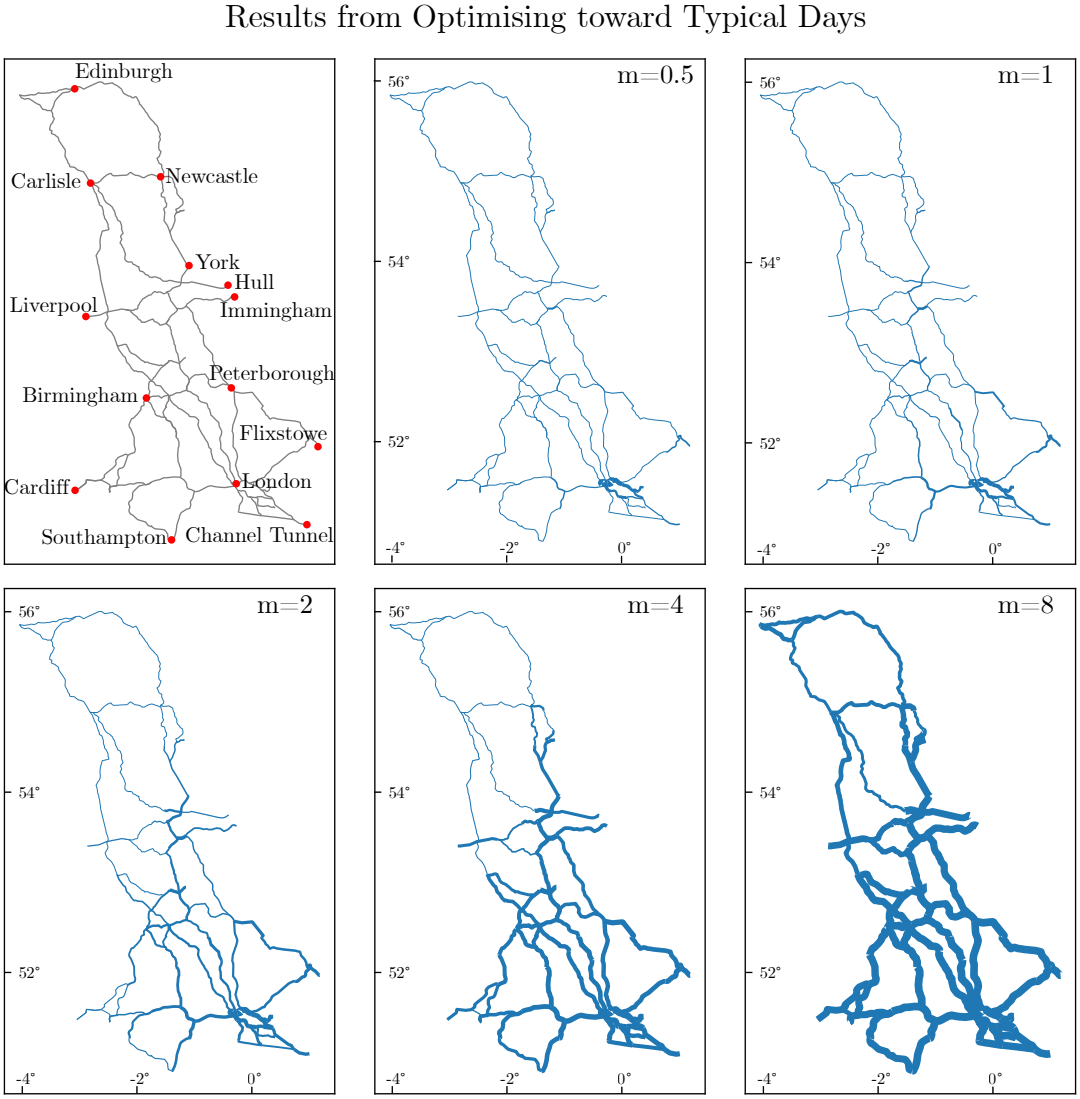


Figure 5.20: Maps showing the resource allocation solution obtained from optimising toward the 250 representative examples days selected from the summer days of the year 2051-2100. Width of the lines indicates the increment in design load $\Delta\mu_i$ for the asset. m is the variable used to quantify the amount of investment. $m = 2$ means the amount of investment equals the amount required to increase the design load for all assets in the network by 2°C.

by the metric based method can only approximate the objective function minimised by the numerical method to a certain degree. Though the performance of the PSO can be improved through parameter tuning, this research project does not intend to study the tuning of those parameters, and there is already a substantial amount of literature available on the subject (e.g. (Oldewage et al., 2020; Bonyadi and Michalewicz, 2016)). The struggles with implementing the PSO algorithm to an optimisation problem formulated with an infrastructure system model of 84 edges suggest that meta-heuristic techniques may face difficulties when dealing with large-scale infrastructure resource allocation problems.

A few simplifications and assumptions are made in the system model construction and specifications of input variables. This case study on rail freight network shares the same limitations of lacking asset condition information as the rail passenger case study as discussed in Section 3.3.6. Furthermore, the setup of investment levels $m \in [0.5, 1, 2, 4, 8]$ and the upper boundary of 10°C are both arbitrary. Moreover, the service disruption and investment resources are not of comparable measurements. If the service disruption and investment resources can both be expressed in monetary units, it is possible to alternate the current single objective optimisation model into a multi-objective one, where the objectives are set as to minimise the cost and service disruption. In this way,an optimum point in investing in capital renewal of the infrastructure system could potentially be found.

5.4 Chapter Conclusion

Case study 1 used a simple network model, the optimisation problem of which, can be solved analytically and numerically to an exact solution. Results from case study 1 point to the fact that the metric-based method works better with smaller amounts of resource inputs and promotes the development of an iterative version of the metric-based method. Using a collection of network models of 3 to 6 nodes, the optimisation problems of which, can still be solved numerically with SQP, case study 2 confirms that the iterative metric-based method can attain a solution nearly as good as the numerical method. Finally, the practicality of the proposed resource allocation approach is demonstrated through its application to the UK's core freight network. This case study showcases the proposed method's ability to effectively handle resource allocation problems for large-scale infrastructure systems as the computational complexity does not increase exponentially with the number of dimensions.

The case study also provides some useful insights. Using the Gini coefficient, it was found that the inequality in asset upgrades is higher with more limited resources. The implication is that when resources are scarce, from a pure perspective of minimising system-level service disruption, it may be more advantageous to focus them on a smaller number of more important assets. Through the correlation analysis, it was found that the relative cost of upgrading an edge and the amount of traffic passes through an edge dominates the resource allocation equally and more than the relative probability of failure. Though the probability of failure does not show a comparable level of correlation as the cost and the flow throughput, it still plays a role

in the resource allocation process and signifies the necessary of integrating the failure patterns and the probability of failures into the optimisation framework.

Chapter 6

Discussion

6.1 Chapter Introduction

In the previous chapters, methods have been proposed to address the research questions asked at the beginning of this thesis and applied to case studies. A systematic climate resilience assessment framework for infrastructure systems was proposed and illustrated with a case study in Chapter 3. A novel network metric that can efficiently direct resource allocation for infrastructure assets was identified in Chapter 4, which is then further refined, made specific and applied to three case studies in Chapter 5. The outcomes and insights from the case studies were already briefly discussed in Chapters 3 and 5. This Chapter instead presents a wider discussion of the method proposed in this thesis. In particular, the following points will be discussed:

- How the proposed methods addresses Infrastructure network systems resilience to climate change.
- The limitations of the proposed methods.
- Challenges in implementing the proposed methods as observed from the case studies.
- The applicability of the proposed method to infrastructure systems in general.

[1]

[1]The current chapter contains materials that were previously prepared for the following: Li, Q., Punzo, G., Robson, C., Arbabi, H. and Mayfield, M., 2022. A Systematic Approach to Climate Resilience Assessment of Infrastructure Networks

6.2 Addressing Infrastructure Network Resilience to Climate Change

This section focuses on discussing how the proposed methods answer the research questions asked at the beginning of this thesis and its novel contribution to climate resilience studies of infrastructure systems. Providing the ongoing climate change and corresponding challenges faced by infrastructure systems, the first research question asked is:

> *How can the magnitude of future infrastructure service disruption due to climate change be estimated accounting for corresponding uncertainties?*

The concept of resilience and resilience assessments are identified as a suitable framework to address the research question asked. Through the literature review, inadequacies in existing resilience assessment approaches were found and three research objectives were set. The following discussion focuses on reflecting on whether the work presented in this thesis fulfils the research objectives set, with the novelty of the work and its contribution to knowledge.

A resilience assessment framework that uses standard climate model output data is developed in Section 3.2. Where the contributing climate variables of a given climate hazard to an infrastructure system can be listed, the future projections for the identified climate variables can be downloaded. By projecting the spatial data onto the geographical shapefile of the physical assets of the infrastructure system, the probability of failure is calculated as a function of the values of the climate variable (weather parameter). This probability of failure is then used to simulate system-level failure scenarios using the Monte Carlo simulation. Through the use of the fragility functions and the Monte Carlo simulation under a resilience assessment framework, the proposed method allows for the estimation of service disruptions in infrastructure systems due to climate change.

The method marks significant progress from a mere count of the number of days that a certain threshold is exceeded under various climate change scenarios, e.g. Sanchis et al. (2020), or the estimation of the number of assets under a qualitative level of risk, e.g. Betts et al. (2021). The proposed approach progresses climate change impact assessment from component-level asset failure-based impact quantification to system-level service disruption-based quantification. The case study results also demonstrate the importance of this progress. The amount of asset failure does not scale linearly with the rising temperature. As shown in Section 3.3.4 Figure 3.9(a), there is an threshold temperature (25°C in this model) after which there is an surge of asset failure. The extent of initial and cumulative service loss also does not scale linearly with the number of asset failures as shown in Figure 3.9(d). Climate change impact assessments should continue beyond the component or physical damage level, which is not a true reflection of the forthcoming disruptions.

The direct use of climate change projection data for probabilistic asset failure modelling offers advances compared to the most advanced weather generation methods used in literature (Panteli and Mancarella, 2015a; Panteli et al., 2016; Fu et al., 2018). Those methods involve

some form of manipulation of present-day weather statistics, where the underlying assumption presumes that future weather patterns remain the same as the current day. Given that climate change causes changes in the temporal and spatial distributions of climate variables (Seneviratne et al., 2021), future weather patterns could be differ from the current day weather patterns spatially. Therefore, methods mentioned above could fall short of simulating representative future failure scenarios. The proposed method here overcomes such shortcoming and can generate behaviourally realistic and physically viable climate hazards under future climate change scenarios. The necessity of using realistic and representative future weather patterns for failure scenario modelling in resilience assessment comes from the understanding that infrastructure systems as complex networks are likely to behave differently under different failure scenarios (Berche et al., 2009). The inference of such understanding is that the resilience of the infrastructure systems to present-day weather patterns can differ from the future climate.

Further advantages offered by the proposed method is that it is possible to estimate the climate resilience of infrastructure systems accounting for the uncertainties in the climate change future evolution. With the proposed method, a wide variety of datasets about future climate projections can be exploited as they get refined by the prolific research area looking at climate change from an atmospheric physics angle. One climate model output can only provide one possible future outcome, and only an extensive collection of outputs from different climate model simulation runs can give a more thorough prediction with the degree of uncertainty and error range addressed. Likewise, a more informative and reliable conclusion regarding the impact of climate change is one that is drawn through the evaluation of assessed resilience to adequate sets of climate model outputs from different global and regional models with different model ensembles and initiation. With thorough implementation of the proposed method, it is possible to convert the climate change predictions to service disruption predictions.

Given the proposed climate resilience assessment framework being able to estimate the magnitude of the future infrastructure service disruption due to climate change, it naturally follows the question of how this disruption can be minimised.

> *How should resources be allocated in the most efficient manner to upgrade infrastructure assets so that the forthcoming service disruptions can be minimised?*

To address such a question, an optimisation problem is first formulated under the resilience assessment framework developed in Chapter 3. For a given weather profile, the assets' local weather parameters are firstly converted to failure probabilities through the fragility functions. The asset failure probabilities are then converted to OD path failure probabilities through the link-route incidence matrix. Multiplying this route failure probability with the route's normal service level gives the expected service disruption on this OD pair. The system level service disruption is estimated by adding the service disruptions of all pairs, which is set as the objective function of the optimisation problem. This service disruption refers to the initial service disruption at the onset of the extreme event and is not the cumulative service disruption from the time of disruption occurring to the point. By differentiating the objective function with respect to an asset's current failure load, the local gradient, or cost/benefit of investing in an

edge, is found. This gradient is regarded as a network metric. After the gradients of all assets are found, the available resources are allocated in proportion to the metrics calculated.

The proposed network metric encapsulates the topology, service and failure probability of assets in one single measurement to indicate the relative priority of asset upgrade. Metrics from the reviewed literature tend to be centred around the topological aspect leaving the supply-demand feature and failure probability out of consideration. Even when those aspects are considered, there is no clear rationale behind explaining why the componential metrics are added or multiplied in such a way. For example, the adding up of a topological importance measure and a functional importance measure in Fujita et al. (2018). Another example is the operation of multiplying the original amount of flow through a node with the betweenness centrality of a node divided by the node degree as a node importance measure in Dunn and Wilkinson (2013). Derived from the differential of the objective function, the metric proposed here comprehensively synthesises the three critical aspects of asset features into one measurement systematically, offering great advantages compared to the ones in the reviewed literature.

The calculation of network metrics and the computations of resource allocation by the metrics are of comparatively low complexity compared to the solution-finding algorithms for high-dimensional optimisation problems. The complexity here refers to the amount of computational resources required to solve a given task, such as time, number of processor cycles and memory space (Dean, 2021). The metric-based method is of lower complexity as the 'solution' can be computed using a modern computer in a fixed number of steps. The steps are described in Section 4.3.3 to calculate the metrics and the steps in Section **??** to compute the $\Delta\mu$ vector. In contrast, when the same formulated problem is regarded as a multivariate nonlinear programming problem with equality and non-equality constraints, the problem is NP-hard (non-deterministic polynomial time hard) (Dean, 2021). As the case study on GB's core freight rail shows, the time taken for a meta-heuristic algorithm to find a local minima for an 84-dimension optimisation problem is a matter of hours. While the time taken for the approach using the network metrics to compute the resource allocation vector is a matter of minutes. The low complexity and the fixed number of steps of the metric-based method mean it scales well as the number of dimensions increases and is capable of handling similar optimisation problems for large-scale infrastructure systems.

The proposed iterative method can allocate available resources efficiently. In the case study 2 (Section 5.2), where the optimisation problems are still solvable with the minimisation functions of a conventional Python package, the proposed iterative metric-based method failed to minimise the objective function to the same degree as the numerical method. However, the gap between the two methods decreases with the increasing number of iterations. After five iterations, the gap can be considered negligible compared to the improvement in service disruption (Figure 5.7). In case study 3, with the core rail freight network having 84 edges, the minimisation function of the Python package failed to yield solutions, and therefore a particle swam algorithm was used. This time, the proposed iterative metric-based method outperforms the numerical method and consistently yields better resource allocation solutions with smaller objective functions.

The work presented in this thesis can answer the research questions asked and advances methods and approaches in the reviewed literature. The developed resilience assessment with the proposed failure initiation method provides a feasible approach to estimate the magnitude of infrastructure service disruption due to climate change. The proposed iterative metric-based resource allocation strategy is able to allocate resources efficiently for large-scale infrastructures to minimise potential future disruptions due to climate change.

6.3 Limitations of the Proposed Approach

With the proposed framework and methods focusing on addressing the geographical variations between assets, their ability to address issues related to the temporal dimension is limited. In recreating weather events and climate hazards for infrastructure resilience assessment purposes, spatial patterns and temporal patterns are two major aspects which are of equal importance and interest. A large portion of the reviewed literature that has undertaken the time series analysis either assumes homogeneous weather conditions across the entire network or randomly samples weather conditions. This work intends to focus on improving the resolution of the spatial features of climate hazards. Therefore more emphasis is put on embedding the spatial pattern into the simulation through the use of realistic weather profiles generated by General Circulation Models (GCM), and it is the first of its kind to utilise climate projection data directly for failure generation in a resilience assessment framework.

This decision is made primarily because the computational power available simply can not support the growth of the scenario tree when the framework features both probabilistic failures with Monte Carlo simulations and time-series analysis. The current sample size for the Monte Carlo simulation is 250. If one more disruptive event is introduced before full recovery and uses the same sample size of 250, the number of possible outcomes will become 250*250 = 62500. The scenario tree grows exponentially. In the current setup, simulations regarding each example day are submitted as a batch job carried out by an Intel Xeon E5-2630 v3 CPUs with 16G of RAM. The CPU time for the 285 example days with climate-based failure scenarios is already 1792 hours in total. With parallelisation, the wall-clock time is about 100 hours. It is impractical to allocate the computational power required to simulate more failures. If such an amount of computational power do become available, or a much smaller network model that requires less RAM, it would be really worth carrying out through time-series analysis, which considers further failures occurring when the percentage of satisfied demand $Q(t)$is not 100% that is, when the network is already disrupted by some previous events.

Due to the impracticality of a time-series approach, the case study on the rail passenger network treats the weather condition across the country on a single day as an instance of situations the railway system might face and does not consider the effects of continuous hot days or heatwaves even though hot weather usually lasts for several days in reality. This assumption is acceptable in this case study because, in real-world practice, repair work will only be carried out once the heatwave is passed (Network-Rail, a). By the operational standards from Network Rail,

track maintenance work should not be carried out when the rail temperatures are above $32°C$ (corresponding to an air temperature of $21°C$) or are predicted to exceed $38°C$ (corresponding to an air temperature of $25°C$) within three days of work conducted. Therefore, an individual day is seen as a single extreme weather event that lasts 24h and uses the maximum near-surface air temperature of the period to calculate the probability of failure. Heatwaves lasting more than one day (24h) can be treated as a single extreme event, with repair work only happening after the heatwave. Other approaches, implying less restrictive assumptions, might be explored, but this was deemed part of possible further refinements. Within the scope of this work, the proposed approach is functional and fulfils its purpose of enable the large scale analysis needed here.

A limitation of the proposed metric-based method is that it is capable of answering 'where' to allocate resources but lacks the ability to answer 'when'. It works for snapshots of infrastructure systems with a given set of asset condition vector $\boldsymbol{\mu}$, route-link incidence matrix $\boldsymbol{H}$, OD flow vector $\boldsymbol{F}$, unit cost vector $\boldsymbol{c}$, and given amount of investment C, under a given snapshot of weather profile $\boldsymbol{\omega}$. With the above variables specified, the proposed method can compute a solution vector $\Delta\boldsymbol{\mu}$ that uses the invested resources efficiently to minimise the expected service disruption. The solution vector $\Delta\boldsymbol{\mu}$ answers which asset to invest and by how much. There is no consideration of time. The missing dimension of time is a limitation as infrastructure systems are constantly evolving, and physical infrastructure assets never stop ageing and deteriorating. Further, the execution of asset upgrade plans is a matter of years, during which changes could happen to the infrastructure systems.

Literature in asset management problems often uses prediction models to factor in future performance and uncertainties into the optimisation model (Marzouk and Omar, 2013; Rashedi and Hegazy, 2015; Hosseini and Smadi, 2021). These prediction models use various statistical techniques to predict the conditions of assets in the future. The lifetime of infrastructure assets can firstly be divided into multiple control periods. At each control period, the optimisation model can be set to optimise the overall asset condition or health of the network accounting for all assets in the network and their predicted conditions. In this way, it adds the temporal dimensions to the optimisation model. The combination of prediction models and the proposed resource allocation method points to a potential future research direction.

As the proposed method is unable to consider continuous time series analysis, the proposed resource allocation approach is applied to two collections of daily profiles, containing top 5 hottest days each year and the 251 representative example days for 2051-2100. Each daily profile is treated as a scenario or weather pattern the railway system might face. For a given investment level, a solution is generated for each daily profile in the collection of days considered. The solutions for all days in a collection are then averaged along each edge to give a combined solution. It must be noted that the optimal solution to a single day may not be the optimal solution for another day, and the combined 'optimal' solution may not be the optimal solution for any day.

In summary, both the climate resilience assessment framework and the resource allocation strat-

egy proposed have weaknesses in accounting for the system's dynamic in the temporal dimension. Yet they represent clear step forward with respect to the current state of the art as emerging from the literature. Further development of the proposed methods to include the temporal dimension should consider the balance between spatial and temporal variation.

6.4 Challenges in Implementing the Framework

The spatial and temporal resolution of the climate change data could be a potential limitation for future applications of the proposed method, depending on the types of climatic hazards of interest. The highest spatial resolution offered by EURO-CORDEX is ∼12km, and the majority of the outputs are with a daily temporal resolution. Although the case studies here only require daily weather profiles, future applications of the method may require weather profiles of higher temporal resolution. The uncertainty in climate model output resolution might be a challenge in some cases but shouldn't stop the application of the developed failure generation method for assessing the climate resilience of infrastructure systems. Resilience of infrastructure systems to the future climate should be studied with future weather patterns. The use of current-day weather profiles could lead to wrong conclusions. It is better to be roughly right than precisely wrong. The extensive number of climate model output datasets need to be better utilised to offer more comprehensive projections of disruptions induced by climate change on infrastructure systems.

The biggest challenge shared by case studies of both the climate resilience assessment framework and the resource allocation strategy is the lack of empirical data and information on industrial practice, which largely affected the representativeness and quality of the infrastructure model constructed. This quality predominantly refers to the quality of asset information data and the assumptions made for the network dynamics.

Due to a lack of asset information, in the case studies, all track segments are assumed to have the same fragile function with the same critical temperature. In reality, the Critical Rail Temperature (CRT) varies from track to track and depends on many factors, including the stress-free temperature of the rail, the quality and degree of consolidation of the ballast, ballast shoulders, tamping, the type and condition of the sleepers and fastenings, and maintenance and renewal work (Network-Rail, b). For the UK's railway network, Network Rail has databases for the CRT of all tracks. Nevertheless, by the time this work was done, such information was not available. Provided such information becomes available, which informs asset-specific fragility functions, more rigorous quantification of infrastructure resilience to future climate can be addressed and more employable resource allocation strategy can be developed.

Information on system dynamics refers to the ways the system responds post-disaster, for example, the industrial practice of rerouting, the prioritization of repairing, and the estimation of spare capacity. A few assumptions are made in the case study of the climate resilience assessment framework to the most general form to complete the resilience assessment setup in

the absence of sufficient information. Improvements can be made to better reflect the reality. The current rerouting algorithm for passenger flows is an adapted Edmond-Karp maximum flow minimum cost algorithm, which all passengers can find and travels on the shortest path. Improvement can be made in adapting transportation flow forecasting models developed in the literature. Advancements can also be made to the recovery processes. The probability of repair at the moment is kept at a 50% chance for all damaged assets. Future works can introduce a 'cap' to limit the number of assets that can be repaired at a given time-step to reflect the limited resources to perform system repair. Alternatively, express the probability of recovery as a function of multiple factors, such as the geographical proximity to repairing resource centres and the next day's weather conditions. If the probability of recovery is set to be associated with topological network metrics, e.g. betweenness centrality, different network recovery strategies can be tested with the proposed framework.

In summary, as the conclusions from implementing the proposed approach are hugely influenced by the quality of the infrastructure model constructed, it is vital to improve the representativeness of the models. In the case study, simplifications and assumptions are made to a few aspects of the assessment framework so that it can demonstrate the proposed network disruption initiation approach and probabilistic asset failure modelling.

6.5 Applicability to General Infrastructure Systems and Weather Hazards

The proposed failure initiation approach is considered applicable to any transport infrastructure system with most types of climate hazards, provided the fragility functions become available with a working infrastructure model. An infrastructure system, whether a water distribution system or a power distribution system, comprises of a collection of physical assets and relies on those assets to provide services. Those assets are designed and operated under specified standards. Weather extremes cause system disruptions when those standards are exceeded and assets malfunction. As long the chain of relationship from the weather parameters of a climate hazard to asset failure, then to the service loss can be mapped and expressed clearly, projected future weather profile data can be employed to simulate disruptive events and estimate the impact of climate change on the system under study.

It is worth noting that though the interdependent nature of the infrastructure systems was not explicitly addressed in all of the case studies, the proposed methods are capable of addressing it. The climate resilience assessment framework developed uses a network model that has one service layer and one separated asset layer that the service layer is dependent on. This choice of one asset layer only was made in the spirit of keeping the system models as simple as possible but no simpler. As the system model here features separate asset and service layers, the climate resilience assessment framework and the resource allocation approached build under the framework can integrate infrastructure interdependence by introducing extra asset layers.

Furthermore, the effects of cascading failure is important to some infrastructure networks, particularly power networks. Cascading failures can be integrated into the framework by introducing an extra iteration of failure determination. In the current assessment framework, the local weather parameter is converted to a probabilistic of asset failure, which is then used to sample asset failure scenarios. Consider the failed assets in those failure scenarios as 'primary failures' caused by weather extremes directly. Cascaded asset failures can be modelled as 'secondary failures' as a consequence of the primary asset failures. Service disruption can be estimated after the asset failures are settled or alongside the chain of cascaded asset failure. The integration of cascading failure into the resource allocation strategy can be achieved by introducing an extra matrix that details the cascading relations between assets. However, estimating how much this would further complicate the optimisation problem is challenging.

All case studies in this work use the same high temperature-related track buckling as the example hazard. In this sense, they deal with one climate hazard (e.g., heat-induced track buckling) associated to a single climate variable (e.g., air temperature). With the proposed approach, it is possible to consider compound weather events that are associated with multiple climate variables. Given a compound event of interest for a specific infrastructure sector, such events can be reconstructed from climate model outputted climate variables (Zscheischler et al., 2020). It provides a viable way to draw an understanding of future compound events, in particular compound events with new combinations of type, timing, severity and locations.

In summary, by introducing multiple inter-dependent asset layers subjected to multiple climate hazards, the proposed approach has the potential to simulate cascading failures in an interconnected infrastructure network to assess and enhance the resilience of the system to single and compound climate hazards.

The proposed metric-based resource allocation strategy, in essence, is a cost/benefit analysis. As concluded in the literature review, metric-based importance ranking works best when tailored to the system under study and reflects the system's dynamics. The proposed metric is applicable to any infrastructure system that shares the same system dynamics as the system model used here. To which extent the proposed method is applicable to other systems remains an open question. However, the methodological approach that resulted in the identification of the metric could be followed and applied to systems of different dynamics, which could potentially lead to the proposal of new metrics.

6.6 Chapter Conclusion

This chapter presents a discussion on the novelty, limitations, implementation challenges and general applicability of the developed climate resilience assessment framework and the proposed resource allocation strategy using a novel network metric. Though the proposed methods have limitations and face many challenges in implementation, they represent a clear step forward with respect to the state of the art. Moreover, they have the potential to develop into a useful

tool to comprehensively assess the climate change impacts on infrastructure systems and aid long-term infrastructure investment decision-making. Alongside the discussion, some potential future developments are briefly mentioned, which will be summarised in the next chapter.

Chapter 7

Conclusion

7.1 Main Contributions and Findings

Recognising the challenges infrastructure systems face due to the changing climate, this work has focused on delivering novel approaches that integrate the changing climate and weather patterns into resilience assessment frameworks and resource allocation strategies.

The first research question asked is:

> *How can the magnitude of future infrastructure service disruption due to climate change be estimated accounting for corresponding uncertainties?*

The reviewed literature in Chapter 2 identifies a research gap between climate change impact assessments and resilience assessments of infrastructure systems. On the climate change impact assessments side of the gap, existing works that engage with climate models remain narrow in focus in dealing only with asset-level or component-level physical damage risk assessments, e.g. counting the number of days passing a certain threshold. On the infrastructure resilience assessment side of the gap, existing works are totally inadequate in bringing behaviourally realistic climate change-induced future disruptions into the resilience assessment framework and accounting for the associated geographical variations in asset failure probabilities (Objective 1).

In Chapter 3, a method to generate network failure scenarios and system disruptions using climate change research data is first presented. Using fragility functions that transfer local weather parameters into probabilities of asset failure, it converts geographical weather data into spatially explicit failure probabilities, which are then used for failure scenario sampling. The proposed method is integrated into a resilience assessment framework to assess system-level functional loss based on the service level retained, which considers possible reconfiguration, as opposed to a mere count of the failed nodes or the identification of a threshold for network fragmentation (Objective 2).

Following the development of the systematic climate resilience assessment framework, a case study is carried out to showcase the proposed method. It attempts to quantify the resilience of

Great Britain's railway passenger transport system to high-temperature-related track buckling under the RCP8.5 climate change scenario over a 95-year horizon. Even though the case study carried out is more of an experimental prototype of the proposed framework, it demonstrates how the developed climate resilience assessment framework is capable of addressing the first research question asked (Objective 4).

Moreover, findings from the case study support the two arguments that motivate this work: 1) Random failure models tend to overestimate the network's resilience and the railway passenger transportation network response differently to different attack strategies; 2) The scale of the system-level service disruption degrades non-linearly with the magnitude of the disruption and the magnitude of disruption scale non-linearly with increasing temperature. Together, they prove the need for bridging the gap between climate change impact assessments and resilience assessments of infrastructure systems.

The second research question asked is:

> *How can the magnitude of future infrastructure service disruption due to climate change be estimated accounting for corresponding uncertainties*

In Chapter 2, two clusters of candidate methods for the above-stated question are reviewed, and a gap is identified. The network metric approach requires a relatively small amount of computational resources but is limited by its disregard for the supply-demand features and probability of component failure. The optimisation model approach is able to optimise toward network-level service disruption but often turns out to be high-dimensional problems that are complicated to solve. Considering the two methods in fact have the same goal of minimising the given objective function, a certain degree of consistency and agreement should exist between them and links the two fields together (Objective 1).

To this end, Chapter 4 first formulated an optimisation model that aims to minimise the service disruption of an infrastructure system model comprising a service layer and an asset layer, constrained by a given amount of resources. The formulated problem is approached using both the network metric method and the optimisation method. Differentiating the objective function yields a network metric that encapsulates the supply-demand feature and the asset-level failure probabilities. The generalised analytical solution for a 2D version of the optimisation problem returns a solution that contains the same expression as the network metric. This discovery bridges the gap for the model in this work and shows that the gap identified above is not impossible to bridge (Objective 3).

Chapter 5 uses the above-discovered network metric to calculate resource allocation solutions for a core freight rail transport network. The metric-based method outperforms a conventional numerical method for optimisation problems in both objective function values and computational time. In the case study, optimisations are performed towards a collection of extremely hot days and also example representative days selected from climate model outputs. Up to this point, it can be concluded that the proposed iterative metric-based resource allocation strategy is able to allocate resources efficiently for large-scale infrastructures to minimise potential future

disruptions due to climate change (Objective 4).

The case study findings point out that the solution varies with the amount of budget allocated. A low level of investment leads to higher inequality in asset upgrades. It implies that when resources are limited, concentrating them on fewer but more critical assets could be more beneficial than spreading the resources out. Most importantly, it was found that the variation in climate-related failure probability plays an indispensable role in allocating resources in the most efficient possible way.

This work centred on systematically quantifying and mitigating the impacts of climate change on critical infrastructure. The proposed methods and the findings from the case studies advance the existing understanding of the climate resilience of infrastructure systems and strategies on resource utilisation for climate change adaptation. Their implementation is imperative in avoiding the worst consequences of climate change.

7.2 Recommendations for Future Works

Based on the case study discussions in Section 3.3.6 and Section 5.3.9, and discussions on the proposed methods themselves in Chapter 6, some recommendations are made for future research projects.

- **Time series analysis** -This work focuses on embedding the spatial pattern into the simulation through the use of realistic weather profiles generated by General Circulation Models (GCM). Future work could further develop the assessment framework to incorporate through time series analysis to better understand the effects of climate change-included changes in the spatial and temporal distributions of climate variables. The effects of prolonged extreme events, such as continuous heatwaves or heavy rains, can be quantified. During extended periods of high temperatures, unresolved asset failures from preceding days could lead to more severe and extensive service disruptions. Antecedent moisture condition from prior rainfalls is an important factor in the occurrence of flood events.
- **Investigate the cause of difference between random attacks and climatic hazards** - It was hypothesised that when subjected to climate hazards, the assets are of unequal probability of failure, and those of close geographical proximity are likely to have a similar probability of failure. While in random attack scenarios, all assets are of equal failure probability. Future works could examine if such spatial correlations of weather-induced failures lead to clusters of assets within close geographic proximity being affected simultaneously and if this bulk removal of assets within close geo-proximity causes the differences between random attacks and climate hazards.
- **The balance between resolution and computational cost** - The railway passenger model used in the case study featured a total of 2,282,270 OD flows, which incurred significant computation costs. Consequently, the simulations were reduced to a set of repre-

sentative example days, instead of every day to 2100. A higher resolution of infrastructure and weather profiles can indeed provide more accurate results. However, a delicate balance between resolution and computational cost is critical and future research should carefully consider the trade-offs involved.

- **Pre- and post-disaster investment trade-off** - Investing in infrastructure system capital renewal is one of many ways to adapt to the future climate. An alternative way is to invest in building up recoverability as the cumulative service disruption is a combined result of the magnitude of initial onset service disruption and speed of asset recovery. Future work could study the trade-off between investing in pre-disaster system robustness or post-disaster recoverability.

- **Prediction model** - The proposed resource allocation method is capable of answering 'where' to allocate resources but can not answer 'when'. Literature in asset management problems often uses prediction models for assets to factor in future performance and uncertainties into the optimisation model. There also exists a research field concerns decision making in deep uncertainty and dynamic planning. Future work could explore the possibility of combining those research efforts with the optimisation model formulated here and the potential changes to the proposed metric.

- **Infrastructure systems beyond rail and transport** - So far in this work, the railway transport systems in the case studies are of the same network dynamic. Future work should see the application of the proposed failure initiation method to resilience assessments of infrastructure networks of different natures. Future work could also follow the trajectory of the journey that leads to the discovery of the metric in this work with infrastructure systems of different dynamics, potentially leading to the discovery of new metrics.

- **Infrastructure interdependency** - Even though the proposed methods are capable of considering infrastructure system interdependence, the case studies only use system models of only one asset layer. Future work could apply the proposed methods to system models with more than one asset layer.

- **Better infrastructure model and data** - Infrastructure owners serve as the primary end users for the climate resilience assessments and optimization models in this work and future works. Simultaneously, they play a vital role in constructing the infrastructure models and specifying the system dynamics for simulating asset failures and system responses in the assessment and optimisation models. Enhancing the collection of asset information data and operational data (including delays and their attribution) for better infrastructure models would be advantageous for infrastructure owners themselves.

All models are wrong, but some are useful.

— George Box

Bibliography

R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *nature*, 406(6794):378–382, 2000.

G. Augusti, A. Borri, and M. Ciampoli. Optimal allocation of resources in reduction of the seismic risk of highway networks. *Engineering Structures*, 16(7):485–497, 1994.

N. Y. Aydin, H. S. Duzgun, H. R. Heinimann, F. Wenzel, and K. R. Gnyawali. Framework for improving the resilience and recovery of transportation networks under geohazard risks. *International journal of disaster risk reduction*, 31:832–843, 2018.

M. Bababeik, N. Khademi, A. Chen, and M. M. Nasiri. Vulnerability analysis of railway networks in case of multi-link blockage. *Transportation Research Procedia*, 22:275–284, 2017.

K. Bakalis and D. Vamvatsikos. Seismic fragility functions via nonlinear response history analysis. *Journal of structural engineering*, 144(10):04018181, 2018.

J. W. Baker. Efficient analytical fragility function fitting using dynamic structural analysis. *Earthquake Spectra*, 31(1):579–599, 2015.

A.-L. Barabási. Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1987):20120375, 2013.

H. Baroud, K. Barker, J. E. Ramirez-Marquez, et al. Importance measures for inland waterway network resilience. *Transportation research part E: logistics and transportation review*, 62: 55–67, 2014.

E. Barrena, D. Canca, L. C. Coelho, and G. Laporte. Exact formulations and algorithm for the train timetabling problem with dynamic demand. *Computers & Operations Research*, 44: 66–74, 2014.

BBC. Storm arwen - bbc news. `https://www.bbc.co.uk/news/topics/cr65z1rjd1dt`, 2021.

H. Behbahani, S. Nazari, M. J. Kang, and T. Litman. A conceptual framework to formulate transportation network design problem considering social equity criteria. *Transportation research part A: policy and practice*, 125:171–183, 2019.

R. E. Bellman. *Dynamic programming*. Princeton university press, 1957.

B. Berche, C. Von Ferber, T. Holovatch, and Y. Holovatch. Resilience of public transport networks against attacks. *The European Physical Journal B*, 71(1):125–137, 2009.

K. Berdica. An introduction to road vulnerability: what has been done, is done and should be done. *Transport policy*, 9(2):117–127, 2002.

A. R. Berkeley, M. Wallace, and C. Coo. A framework for establishing critical infrastructure resilience goals. *Final report and recommendations by the council, national infrastructure advisory council*, pages 18–21, 2010.

D. P. Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Academic press, 2014.

M. Bessani, J. A. Massignan, R. Z. Fanucchi, M. H. Camillo, J. B. London, A. C. Delbem, and C. D. Maciel. Probabilistic assessment of power distribution systems resilience under extreme weather. *IEEE Systems Journal*, 13(2):1747–1756, 2018.

R. Betts, A. Haward, and K. Pearson. The third uk climate change risk assessment. *Prepared for the Climate Change Committee, London*, 2021.

P. T. Boggs and J. W. Tolle. Sequential quadratic programming. *Acta numerica*, 4:1–51, 1995.

M. R. Bonyadi and Z. Michalewicz. Impacts of coefficients on movement patterns in the particle swarm optimization algorithm. *IEEE Transactions on Evolutionary Computation*, 21(3):378–390, 2016.

M. Bruneau, S. E. Chang, R. T. Eguchi, G. C. Lee, T. D. O'Rourke, A. M. Reinhorn, M. Shinozuka, K. Tierney, W. A. Wallace, and D. Von Winterfeldt. A framework to quantitatively assess and enhance the seismic resilience of communities. *Earthquake spectra*, 19(4):733–752, 2003.

P. Bubeck, L. Dillenardt, L. Alfieri, L. Feyen, A. H. Thieken, and P. Kellermann. Global warming to increase flood risk on european railways. *Climatic Change*, 155(1):19–36, 2019.

F. Cadini, E. Zio, and C.-A. Petrescu. Using centrality measures to rank the importance of the components of a complex network infrastructure. In *Critical Information Infrastructure Security: Third International Workshop, CRITIS 2008, Rome, Italy, October13-15, 2008. Revised Papers 3*, pages 155–167. Springer, 2009.

P. Camus, A. Tomás, G. Díaz-Hernández, B. Rodríguez, C. Izaguirre, and I. Losada. Probabilistic assessment of port operation downtimes under climate change. *Coastal Engineering*, 147:12–24, 2019.

P. Cappanera and M. P. Scaparra. Optimal allocation of protective resources in shortest-path networks. *Transportation Science*, 45(1):64–80, 2011.

P. Chakroborty, P. K. Agarwal, and A. Das. Comprehensive pavement maintenance strategies for road networks through optimal allocation of resources. *Transportation Planning and Technology*, 35(3):317–339, 2012.

R. Chan and J. L. Schofer. Measuring transportation system resilience: Response of rail transit to weather disruptions. *Natural Hazards Review*, 17(1):05015004, 2016.

L. Chen and E. Miller-Hooks. Resilience: an indicator of recovery capability in intermodal freight transport. *Transportation Science*, 46(1):109–123, 2012.

L. Cinquini, D. Crichton, C. Mattmann, J. Harney, G. Shipman, F. Wang, R. Ananthakrishnan, N. Miller, S. Denvil, M. Morgan, et al. The earth system grid federation: An open infrastructure for access to distributed geospatial data. *Future Generation Computer Systems*, 36: 400–417, 2014.

B. Cundill and N. D. Alexander. Sample size calculations for skewed distributions. *BMC medical research methodology*, 15:1–9, 2015.

W. Dean. Computational Complexity Theory. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2021 edition, 2021.

Y. Del Valle, G. K. Venayagamoorthy, S. Mohagheghi, J.-C. Hernandez, and R. G. Harley. Particle swarm optimization: basic concepts, variants and applications in power systems. *IEEE Transactions on evolutionary computation*, 12(2):171–195, 2008.

Deloitte. Assessing the value of rail freight. Technical report, Deloitte LLP, 2021.

T. Ding, S. Liu, W. Yuan, Z. Bie, and B. Zeng. A two-stage robust reactive power optimization considering uncertain wind power integration in active distribution networks. *IEEE Transactions on Sustainable Energy*, 7(1):301–311, 2015.

J. Douris and G. Kim. Wmo atlas of mortality and economic losses from weather, climate and water extremes (1970–2019). Technical report, World Meteorological Organization, 2021.

S. Draper, D. Johnson, A. Kenney, A. Rowe, M. Angus, and M. Thomas. Definitions of railway performance metrics. Technical report, Rail Delivery Group, 2017. URL `https://www.raildeliverygroup.com/about-us/publications/acop/287-rdgntfdorpm-final/file.html`.

S. Dunn and S. M. Wilkinson. Identifying critical components in infrastructure networks using network topology. *Journal of Infrastructure Systems*, 19(2):157–165, 2013.

J. Edmonds and R. M. Karp. Theoretical improvements in algorithmic efficiency for network flow problems. *Journal of the ACM (JACM)*, 19(2):248–264, 1972.

B. Edwards, S. Hofmeyr, G. Stelle, and S. Forrest. Internet topology over time. *arXiv preprint arXiv:1202.3993*, 2012.

J. Elsner and A. Tsonis. Nonlinear prediction, chaos, and noise. *Bulletin of the American Meteorological Society*, 1992.

EU-CIRCLE. EU-CIRCLE: D3.5 holistic CI climate hazard risk assessment framework. URL `https://www.eu-circle.eu/wp-content/uploads/2018/10/D3.5.pdf`.

G. Fagiolo. Clustering in complex directed networks. *Physical Review E*, 76(2):026107, 2007.

Y.-P. Fang and E. Zio. An adaptive robust framework for the optimization of the resilience of interdependent infrastructures under natural hazards. *European Journal of Operational Research*, 276(3):1119–1136, 2019.

R. Faturechi and E. Miller-Hooks. Measuring the performance of transportation infrastructure systems in disasters: A comprehensive review. *Journal of infrastructure systems*, 21(1): 04014025, 2015.

E. Ferranti, L. Chapman, C. Lowe, S. McCulloch, D. Jaroszweski, and A. Quinn. Heat-related failures on southeast england's railway network: Insights and implications for heat risk management. *Weather, Climate, and Society*, 8(2):177–191, 2016.

L. Fiondella, A. Rahman, N. Lownes, and V. V. Basavaraj. Defense of high-speed rail with an evolutionary algorithm guided by game theory. *IEEE Transactions on Reliability*, 65(2): 674–686, 2015.

E. M. Fischer and R. Knutti. Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes. *Nature climate change*, 5(6):560–564, 2015.

D. M. Frangopol. Life-cycle performance, management, and optimisation of structural systems under uncertainty: accomplishments and challenges 1. *Structure and infrastructure Engineering*, 7(6):389–413, 2011.

L. C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41, 1977.

G. Fu, S. Wilkinson, R. J. Dawson, H. J. Fowler, C. Kilsby, M. Panteli, and P. Mancarella. Integrated approach to assess the resilience of future electricity infrastructure networks to climate hazards. *IEEE Systems Journal*, 12(4):3169–3180, 2018.

H. Fujita, A. Gaeta, V. Loia, and F. Orciuoli. Resilience analysis of critical infrastructures: a cognitive approach based on granular computing. *IEEE transactions on cybernetics*, 49(5): 1835–1848, 2018.

A. Gaeta, V. Loia, and F. Orciuoli. A comprehensive model and computational methods to improve situation awareness in intelligence scenarios. *Applied Intelligence*, 51(9):6585–6608, 2021.

L. K. Gallos, R. Cohen, P. Argyrakis, A. Bunde, and S. Havlin. Stability and topology of scale-free networks under attack and defense strategies. *Physical review letters*, 94(18):188701, 2005.

M. T. Gastner and M. E. Newman. The spatial structure of networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 49:247–252, 2006.

H. George-Williams and E. Patelli. A hybrid load flow and event driven simulation approach to multi-state system reliability evaluation. *Reliability Engineering & System Safety*, 152: 351–367, 2016.

C. Gini. *Variabilità e mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche.[Fasc. I.]*. Tipogr. di P. Cuppini, 1912.

O. Giustolisi, L. Ridolfi, and A. Simone. Tailoring centrality metrics for water distribution networks. *Water Resources Research*, 55(3):2348–2369, 2019.

N. Goldbeck, P. Angeloudis, and W. Y. Ochieng. Resilience assessment for interdependent urban infrastructure systems using dynamic network flow models. *Reliability Engineering & System Safety*, 188:62–79, 2019.

A. D. González, L. Dueñas-Osorio, M. Sánchez-Silva, and A. L. Medaglia. The interdependent network design problem for optimal infrastructure system restoration. *Computer-Aided Civil and Infrastructure Engineering*, 31(5):334–350, 2016.

S. Greenham, E. Ferranti, A. Quinn, and K. Drayson. The impact of high temperatures and extreme heat to delays on the london underground rail network: An empirical study. *Meteorological Applications*, 27(3):e1910, 2020.

X. Han and D. M. Frangopol. Life-cycle risk-based optimal maintenance strategy for bridge networks subjected to corrosion and seismic hazards. *Journal of Bridge Engineering*, 28(1): 04022128, 2023.

P. Hande, S. Zhang, and M. Chiang. Distributed rate allocation for inelastic flows. *IEEE/ACM Transactions on Networking*, 15(6):1240–1253, 2007.

E. J. Hearnshaw and M. M. Wilson. A complex network approach to supply chain network theory. *International Journal of Operations & Production Management*, 33(4):442–469, 2013.

B. Hickish, D. I. Fletcher, and R. F. Harrison. A rail network performance metric to capture passenger experience. *Journal of Rail Transport Planning & Management*, 11:100138, 2019.

B. Hickish, D. I. Fletcher, and R. F. Harrison. A methodology to optimise a rail network specification for maximum passenger satisfaction and reduced initial investment. *Journal of Rail Transport Planning & Management*, 21:100279, 2022.

F. S. Hillier and G. J. Lieberman. Introduction to operations research, 2001.

H. M. HM Treasury. National infrastructure strategy. Technical report, 2020.

J. Högdahl, M. Bohlin, and O. Fröidh. A combined simulation-optimization approach for minimizing travel time and delays in railway timetables. *Transportation Research Part B: Methodological*, 126:192–212, 2019.

C. S. Holling. Resilience and stability of ecological systems. *Annual review of ecology and systematics*, 4(1):1–23, 1973.

S. Hosseini, D. Ivanov, and A. Dolgui. Review of quantitative methods for supply chain resilience analysis. *Transportation Research Part E: Logistics and Transportation Review*, 125:285–307, 2019.

S. A. Hosseini and O. Smadi. How prediction accuracy can affect the decision-making process in pavement management system. *Infrastructures*, 6(2):28, 2021.

S.-M. Hosseininasab and S.-N. Shetab-Boushehri. Integration of selecting and scheduling urban road construction projects as a time-dependent discrete network design problem. *European Journal of Operational Research*, 246(3):762–771, 2015.

J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. Johnson. *Climate chage 2001.* 2001.

X. Huang, J. Gao, S. V. Buldyrev, S. Havlin, and H. E. Stanley. Robustness of interdependent networks under targeted attack. *Physical Review E*, 83(6):065101, 2011.

R. Hulse and J. Cain. *Structural mechanics.* Bloomsbury Publishing, 2018.

IPCC. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013. ISBN 978-1-107-66182-0. doi: 10.1017/CBO9781107415324. URL `www.climatechange2013.org`.

D. Jaroszweski, E. Hooper, C. Baker, L. Chapman, and A. Quinn. The impacts of the 28 june 2012 storms on uk road and rail transport. *Meteorological Applications*, 22(3):470–476, 2015.

T. J. C. o. t. N. S. S. JCNSS. Readiness for storms ahead? critical national infrastructure in an age of climate change. Technical report, The House of Lords and the House of Commons, 2022.

J. Johansson and H. Hassel. An approach for modelling interdependent infrastructures in the context of vulnerability analysis. *Reliability Engineering & System Safety*, 95(12):1335–1344, 2010.

D. S. Johnson, J. K. Lenstra, and A. R. Kan. The complexity of the network design problem. *Networks*, 8(4):279–285, 1978.

D. B. Karakoc, K. Barker, and A. D. González. Analyzing the tradeoff between vulnerability and recoverability investments for interdependent infrastructure networks. *Socio-Economic Planning Sciences*, page 101508, 2023.

F. P. Kelly, A. K. Maulloo, and D. K. H. Tan. Rate control for communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research society*, 49(3): 237–252, 1998.

M. Kendall. *Rank Correlation Methods.* Theory and applications of rank order-statistics. Griffin, 1970. ISBN 978-0-85264-199-6. URL `https://books.google.co.uk/books?id=Mm2jjgEACAAJ`.

J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks*, volume 4, pages 1942–1948. IEEE, 1995.

A. Kish, G. Samavedam, et al. Track buckling prevention: theory, safety concepts, and applications. Technical report, John A. Volpe National Transportation Systems Center (US), 2013.

J. P. Kleijnen and J. Wan. Optimization of simulated systems: Optquest and alternatives. *Simulation Modelling Practice and Theory*, 15(3):354–362, 2007.

E. E. Koks, J. Rozenberg, C. Zorn, M. Tariverdi, M. Vousdoukas, S. A. Fraser, J. Hall, and S. Hallegatte. A global multi-hazard risk analysis of road and railway infrastructure assets. *Nature communications*, 10(1):2677, 2019.

J. C. Lam, B. T. Adey, M. Heitzler, J. Hackl, P. Gehl, N. Van Erp, D. d'Ayala, P. van Gelder, and L. Hurni. Stress tests for a road network using fragility functions and functional capacity loss functions. *Reliability Engineering & System Safety*, 173:78–93, 2018.

V. Latora and M. Marchiori. Vulnerability and protection of infrastructure networks. *Physical Review E*, 71(1):015103, 2005.

P. Leviäkangas, A. Tuominen, R. Molarius, J. Schabel, S. Toivonen, J. Keränen, J. Törnqvist, L. Makkonen, A. Vajda, H. Tuomenvirta, et al. Extreme weather impacts on transport systems. 2011.

Q. Li, G. Punzo, C. Robson, H. Arbabi, and M. Mayfield. A novel approach to climate resilience of infrastructure networks. *arXiv preprint arXiv:2211.10132*, 2022.

J. Lin and Y. Ban. Complex network topology of transportation systems. *Transport reviews*, 33 (6):658–685, 2013.

M. Liu and D. M. Frangopol. Balancing connectivity of deteriorating bridge networks and long-term maintenance cost through optimization. *Journal of Bridge Engineering*, 10(4):468–481, 2005.

H. Lobban, Y. Almoghathawi, N. Morshedlou, and K. Barker. Community vulnerability perspective on robust protection planning in interdependent infrastructure networks. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 235(5): 798–813, 2021.

J. Ludvigsen and R. Klæboe. Extreme weather impacts on freight railways in europe. *Natural hazards*, 70:767–787, 2014.

H. B. Mann and D. R. Whitney. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, 18(1):50 – 60, 1947. doi: 10.1214/aoms/1177730491. URL `https://doi.org/10.1214/aoms/1177730491`.

J. R. R. A. Martins and A. Ning. *Engineering Design Optimization*. Cambridge University Press, Jan 2022. ISBN 9781108833417.

M. Marzouk and M. Omar. Multiobjective optimisation algorithm for sewer network rehabilitation. *Structure and Infrastructure Engineering*, 9(11):1094–1102, 2013.

V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. Gomis, et al. Climate change 2021: the physical science basis. *Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change*, 2, 2021.

M. McCarter, K. Barker, J. Johansson, and J. E. Ramirez-Marquez. A bi-objective formulation for robust defense strategies in multi-commodity networks. *Reliability Engineering & System Safety*, 176:154–161, 2018.

Met Office. Past weather events. URL `https://www.metoffice.gov.uk/weather/learn-about/past-uk-weather-events`.

Met Office. Unprecedented extreme heatwave, july 2022, 2022. URL `https://www.metoffice.gov.uk/about-us/press-office/news/weather-and-climate/2022/july-heat-review`.

E. Miller-Hooks, X. Zhang, and R. Faturechi. Measuring and maximizing resilience of freight transportation networks. *Computers & Operations Research*, 39(7):1633–1643, 2012.

J. A. Momoh. *Electric power system applications of optimization*. CRC press, 2017.

A. Mostafavi. A system-of-systems framework for exploratory analysis of climate change impacts on civil infrastructure resilience. *Sustainable and Resilient Infrastructure*, 3(4):175–192, 2018.

A. Nagurney, J. Dong, and D. Zhang. A supply chain network equilibrium model. *Transportation Research Part E: Logistics and Transportation Review*, 38(5):281–303, 2002.

National Infrastructure Commission. Anticipate, react, recover. resilient infrastructure systems. Technical report, National Infrastructure Commission, 2020.

National Weather Service. Past significant weather events (1950s til now). URL `https://www.weather.gov/mob/events`.

Network-Rail. Nr/l2/trk/3011: Continuously welded rail (cwr) track, a.

Network-Rail. Nr/l3/trk/7012: Critical rail temperature management for projects, b.

Network Rail. Freight network study. Technical report, 2017a.

Network Rail. Freight network study - summary document, 2017b.

NetworkX. Atlas of all connected graphs with up to 6 nodes., 2023. URL `https://networkx.org/documentation/stable/auto_examples/graphviz_layout/plot_atlas.html`.

M. E. Newman. A measure of betweenness centrality based on random walks. *Social networks*, 27(1):39–54, 2005.

J. Nocedal and S. J. Wright. *Numerical optimization*. Springer, second edition, 2006.

M. Oehlers and B. Fabian. Graph metrics for network robustness—a survey. *Mathematics*, 9 (8):895, 2021.

M. Ohba and S. Sugimoto. Impacts of climate change on heavy wet snowfall in japan. *Climate Dynamics*, 54(5-6):3151–3164, 2020.

E. T. Oldewage, A. P. Engelbrecht, and C. W. Cleghorn. Movement patterns of a particle swarm in high dimensional spaces. *Information Sciences*, 512:1043–1062, 2020.

ONS. The gini coefficient. `https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/methodologies/theginicoefficient#:~:text=Overview-,The%20most%20widely%20used%20summary%20measure%20of%20inequality%20in%20the,equally%20household%20income%20is%20distributed.`, 2022.

ORR. Freight rail usage and performance, 2023. URL `https://dataportal.orr.gov.uk/statistics/usage/freight-rail-usage-and-performance/`.

M. Ouyang. A mathematical framework to optimize resilience of interdependent critical infrastructure systems under spatially localized attacks. *European Journal of Operational Research*, 262(3):1072–1084, 2017.

J. E. Overland, K. Dethloff, J. A. Francis, R. J. Hall, E. Hanna, S.-J. Kim, J. A. Screen, T. G. Shepherd, and T. Vihma. Nonlinear response of mid-latitude weather to the changing arctic. *Nature Climate Change*, 6(11):992–999, 2016.

A. Pagani, G. Mosquera, A. Alturki, S. Johnson, S. Jarvis, A. Wilson, W. Guo, and L. Varga. Resilience or robustness: identifying topological vulnerabilities in rail networks. *Royal Society open science*, 6(2):181301, 2019.

E. J. Palin, H. E. Thornton, C. T. Mathison, R. E. McCarthy, R. T. Clark, and J. Dora. Future projections of temperature-related climate change impacts on the railway network of great britain. *Climatic Change*, 120(1):71–93, 2013.

E. J. Palin, I. Stipanovic Oslakovic, K. Gavin, and A. Quinn. Implications of climate change for railway infrastructure. *Wiley Interdisciplinary Reviews: Climate Change*, 12(5):e728, 2021.

D. P. Palomar and M. Chiang. A tutorial on decomposition methods for network utility maximization. *IEEE Journal on Selected Areas in Communications*, 24(8):1439–1451, 2006.

R. Pant, K. Barker, F. H. Grant, and T. L. Landers. Interdependent impacts of inoperability at multi-modal transportation container terminals. *Transportation Research Part E: Logistics and Transportation Review*, 47(5):722–737, 2011.

R. Pant, J. W. Hall, and S. P. Blainey. Vulnerability assessment framework for interdependent critical infrastructures: case-study for great britain's rail network. *European Journal of Transportation and Infrastructure Research*, 16(1):174–194, 2016.

M. Panteli and P. Mancarella. Influence of extreme weather and climate change on the resilience of power systems: Impacts and possible mitigation strategies. *Electric Power Systems Research*, 127:259–270, 2015a.

M. Panteli and P. Mancarella. Modeling and evaluating the resilience of critical electrical power infrastructure to extreme weather events. *IEEE Systems Journal*, 11(3):1733–1742, 2015b.

M. Panteli, C. Pickering, S. Wilkinson, R. Dawson, and P. Mancarella. Power system resilience to extreme weather: fragility modeling, probabilistic impact assessment, and adaptation measures. *IEEE Transactions on Power Systems*, 32(5):3747–3757, 2016.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python, 2011.

S. Y. Philip, S. F. Kew, G. J. Van Oldenborgh, F. S. Anslow, S. I. Seneviratne, R. Vautard, D. Coumou, K. L. Ebi, J. Arrighi, R. Singh, et al. Rapid attribution analysis of the extraordinary heatwave on the pacific coast of the us and canada june 2021. *Earth System Dynamics Discussions*, 2021:1–34, 2021.

K. Porter. A beginner's guide to fragility, vulnerability, and risk. *Encyclopedia of earthquake engineering*, 2015:235–260, 2015.

H.-O. Pörtner, D. C. Roberts, H. Adams, C. Adler, P. Aldunce, E. Ali, R. A. Begum, R. Betts, R. B. Kerr, R. Biesbroek, et al. *Climate change 2022: Impacts, adaptation and vulnerability*. IPCC Geneva, Switzerland:, 2022.

G. Punzo, R. Beasley, G. Clarke, N. Holt, S. Jobbins, and M. Mayfield. *Challenges of Complexity and Resilience in Complex Engineering Systems*. 2018. doi: 10.13140/RG.2.2.22281.70248.

M.-T. Puth, M. Neuhäuser, and G. D. Ruxton. Effective use of spearman's and kendall's correlation coefficients for association between two measured traits. *Animal Behaviour*, 102:77–84, 2015.

pyswarm. Particle swarm optimization (pso) with constraint support. `https://github.com/tisimst/pyswarm`, 2015.

V. Ramdas and D. Chapman. Project report cpr2282: Options for capacity measures/metrics, 2017. `https://www.orr.gov.uk/sites/default/files/om/trl_report_options_for_capacity_measures_and_metrics.pdf`.

R. S. Rao, K. Ravindra, K. Satish, and S. Narasimham. Power loss minimization in distribution system using network reconfiguration in the presence of distributed generation. *IEEE transactions on power systems*, 28(1):317–325, 2012.

R. Rashedi and T. Hegazy. Capital renewal optimisation for large-scale infrastructure networks: genetic algorithms versus advanced mathematical tools. *Structure and Infrastructure Engineering*, 11(3):253–262, 2015.

S. Rinaldi, J. Peerenboom, and T. Kelly. Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE Control Systems Magazine*, 21(6):11–25, 2001. doi: 10.1109/37.969131.

C. Robson, S. Barr, A. Ford, and P. James. The structure and behaviour of hierarchical infrastructure networks. *Applied Network Science*, 6(1):1–25, 2021.

M. Rummukainen. Changes in climate and weather extremes in the 21st century. *Wiley Interdisciplinary Reviews: Climate Change*, 3(2):115–129, 2012.

H. Sabouhi, A. Doroudi, M. Fotuhi-Firuzabad, and M. Bashiri. Electrical power system resilience assessment: A comprehensive approach. *IEEE Systems Journal*, 14(2):2643–2652, 2019.

I. V. Sanchis, R. I. Franco, P. M. Fernández, P. S. Zuriaga, and J. B. F. Torres. Risk of increasing temperature due to climate change on high-speed rail network in spain. *Transportation Research Part D: Transport and Environment*, 82:102312, 2020.

J. Saramäki, M. Kivelä, J.-P. Onnela, K. Kaski, and J. Kertesz. Generalizations of the clustering coefficient to weighted complex networks. *Physical Review E*, 75(2):027105, 2007.

A. Savitzky and M. J. E. Golay. Smoothing and differentiation of data by simplified least squares procedures. 36(8):1627–1639. ISSN 0003-2700, 1520-6882. doi: 10.1021/ac60214a047. Python implementation: `https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.savgol_filter.html`.

A. Scherb, L. Garrè, and D. Straub. Reliability and component importance in networks subject to spatially distributed hazards followed by cascading failures. *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg*, 3(2), 2017.

S. Seneviratne, X. Zhang, M. Adnan, W. Badi, C. Dereczynski, A. Di Luca, S. Ghosh, I. Iskandar, J. Kossin, S. Lewis, F. Otto, I. Pinto, M. Satoh, S. M. Vicente-Serrano, M. Wehner, and B. Zhou. Weather and climate extreme events in a changing climate. In *Climate Change 2021: The Physical Science Basis*. Cambridge University Press, 2021.

F. Shadman, S. Sadeghipour, M. Moghavvemi, and R. Saidur. Drought and energy security in key asean countries. *Renewable and Sustainable energy reviews*, 53:50–58, 2016.

V. Silva, S. Akkar, J. Baker, P. Bazzurro, J. M. Castro, H. Crowley, M. Dolsek, C. Galasso, S. Lagomarsino, R. Monteiro, et al. Current challenges and future trends in analytical fragility and vulnerability modeling. *Earthquake Spectra*, 35(4):1927–1952, 2019.

A. Smolyak, O. Levy, I. Vodenska, S. Buldyrev, and S. Havlin. Mitigation of cascading failures in complex networks. *Scientific reports*, 10(1):16124, 2020.

V. L. Spiegler, M. M. Naim, and J. Wikner. A control engineering approach to the assessment of supply chain resilience. *International journal of production research*, 50(21):6162–6187, 2012.

J. Sun, S. Balakrishnan, and Z. Zhang. A resource allocation framework for predisaster resilience management of interdependent infrastructure networks. *Built Environment Project and Asset Management*, 11(2):284–303, 2021.

J. Sutherland and J. Wolf. *Coastal defence vulnerability 2075*. HR Wallingford Limited Wallingford, 2002.

C. Ta, A. V. Goodchild, and K. Pitera. Structuring a definition of resilience for the freight transportation system. *Transportation Research Record*, 2097(1):19–25, 2009.

A. N. Tari, M. S. Sepasian, and M. T. Kenari. Resilience assessment and improvement of distribution networks against extreme weather events. *International Journal of Electrical Power & Energy Systems*, 125:106414, 2021.

The White House. Fact sheet: President biden announces support for the bipartisan infrastructure framework, 2021. URL `https://www.whitehouse.gov/briefing-room/statements-releases/2021/06/24/fact-sheet-president-biden-announces-support-for-the-bipartisan-infrastructure-framework/`.

M. T. Todinov. Analysis and optimization of repairable flow networks with complex topology. *IEEE Transactions on Reliability*, 60(1):111–124, 2011.

D. N. Trakas, M. Panteli, N. D. Hatziargyriou, and P. Mancarella. Spatial risk analysis of power systems resilience during extreme events. *Risk Analysis*, 39(1):195–211, 2019.

H. M. Treasury. Autumn budget and spending review 2021. Technical report, 2021.

R. Tsubaki, J. D. Bricker, K. Ichii, and Y. Kawahara. Development of fragility curves for railway embankment and ballast scour due to overtopping flood flow. *Natural Hazards and Earth System Sciences*, 16(12):2455–2472, 2016.

A. Vajda, H. Tuomenvirta, I. Juga, P. Nurmi, P. Jokinen, and J. Rauhala. Severe weather affecting european transport systems: the identification, classification and frequencies of events. *Natural Hazards*, 72(1):169–188, 2014.

W. F. Vincent. Arctic climate change: Local impacts, global consequences, and policy implications. *The Palgrave handbook of Arctic policy and politics*, pages 507–526, 2020.

V. Vita. Development of a decision-making algorithm for the optimum size and placement of distributed generation units in distribution networks. *Energies*, 10(9):1433, 2017.

C. Von Ferber, T. Holovatch, Y. Holovatch, and V. Palchykov. Public transport networks: empirical analysis and modeling. *The European Physical Journal B*, 68(2):261–275, 2009.

D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684): 440–442, 1998.

S. Wilkinson, S. Dunn, R. Adams, N. Kirchner-Bossi, H. J. Fowler, S. G. Otálora, D. Pritchard, J. Mendes, E. J. Palin, and S. C. Chan. Consequence forecasting: A rational framework for predicting the consequences of approaching storms. *Climate Risk Management*, 35:100412, 2022.

J. Yin, L. Yang, T. Tang, Z. Gao, and B. Ran. Dynamic passenger demand oriented metro train scheduling with energy-efficiency and waiting time minimization: Mixed-integer linear programming approaches. *Transportation Research Part B: Methodological*, 97:182–213, 2017.

A. Younesi, H. Shayeghi, A. Safari, and P. Siano. A quantitative resilience measure framework for power systems against wide-area extreme events. *IEEE Systems Journal*, 15(1):915–922, 2020.

M. D. Zelinka, T. A. Myers, D. T. McCoy, S. Po-Chedley, P. M. Caldwell, P. Ceppi, S. A. Klein, and K. E. Taylor. Causes of higher climate sensitivity in cmip6 models. *Geophysical Research Letters*, 47(1):e2019GL085782, 2020.

C. Zhang, J.-j. Kong, and S. P. Simonovic. Restoration resource allocation model for enhancing resilience of interdependent infrastructure systems. *Safety science*, 102:169–177, 2018.

J. Zhang, E. Modiano, and D. Hay. Enhancing network robustness via shielding. *IEEE/ACM Transactions on Networking*, 25(4):2209–2222, 2017.

X. Zhang, E. Miller-Hooks, and K. Denny. Assessing the role of network topology in transportation network resilience. *Journal of Transport Geography*, 46:35–45, 2015.

E. Zio and R. Piccinelli. Randomized flow model and centrality measure for electrical power transmission network analysis. *Reliability Engineering & System Safety*, 95(4):379–385, 2010.

J. Zscheischler, O. Martius, S. Westra, E. Bevacqua, C. Raymond, R. M. Horton, B. van den Hurk, A. AghaKouchak, A. Jézéquel, M. D. Mahecha, et al. A typology of compound weather and climate events. *Nature reviews earth & environment*, 1(7):333–347, 2020.

F. W. Zwiers, L. V. Alexander, G. C. Hegerl, T. R. Knutson, J. P. Kossin, P. Naveau, N. Nicholls, C. Schär, S. I. Seneviratne, and X. Zhang. Climate extremes: challenges in estimating and understanding recent changes in the frequency and intensity of extreme climate and weather events. *Climate science for serving society: research, modeling and prediction priorities*, pages 339–389, 2013.